

Business  
analytics and  
data visualization

# Airbnb Dataset Analysis

---

FINAL PROJECT

MANYU JIANG

# Airbnb Data Overview

---

## KPIS

- Price
- Number of reviews
- Review recency

## DIMENSIONS FOR ANALYSIS

- Neighborhood group
- Neighborhood
- Room Type
- Minimum nights needed to stay
- Availability of the listing in 365 days

# Data Cleaning

---

## OUTLIERS DETECTOR

- Available day = 0 or > 365
- Minimum nights > Available days

## METHOD

Outlier Detector

×

```
IF [Availability 365] = 0 OR [Availability 365] > 365
OR [Minimum Nights] > [Availability 365]
THEN "Outliers"
Else "Usable Data" END
```

The calculation is valid.

16 Dependencies ▼

Apply

OK

## Summary Report

Distinct count of Id	6,426
Distinct count of Name	6,006
Distinct count of Host Id	1,647
Median Price	\$131
Median Number Of Reviews	2
Median Reviews Per Month	1
Distinct count of Neighbourhood	42
Distinct count of Neighbourhood Group	5
Distinct count of Room Type	3
Median Availability 365	318
Median Minimum Nights	3

### Outlier Detector

- ☐ (All)
- ☐ Outliers
- ☒ Usable Data

# Key Facts

---



**6,426** Valid Listings



**5** Neighborhood



**\$131** Median Price



**42** Neighborhood Groups



**2** Median Number  
of Reviews



**3** Room Types

# Interesting Patterns

---

- Which Neighborhood Group is Expensive?
- Which Neighborhood is the Most Expensive in each Neighborhood Group?
- What is the Most Popular Room Type?
- What is the Most Popular Room Type in Central Region?
- Which Neighborhood Has the Most Listings?
- What is the Price Range?

# What Factors May Correlate with Price?

---

- Do Apartments with More Reviews Tend to Have Higher Price Versus the other Groups?
- Do Apartments with More Availabilities Tend to Have Higher Price Versus the other Groups?
- Do Apartments with Less Min Nights Tend to Have Higher Price Versus the other Groups?

# What Factors are Impacting the Price?

---

- What are the Key Factors Impacting the Price?



# Listing Segmentation

---

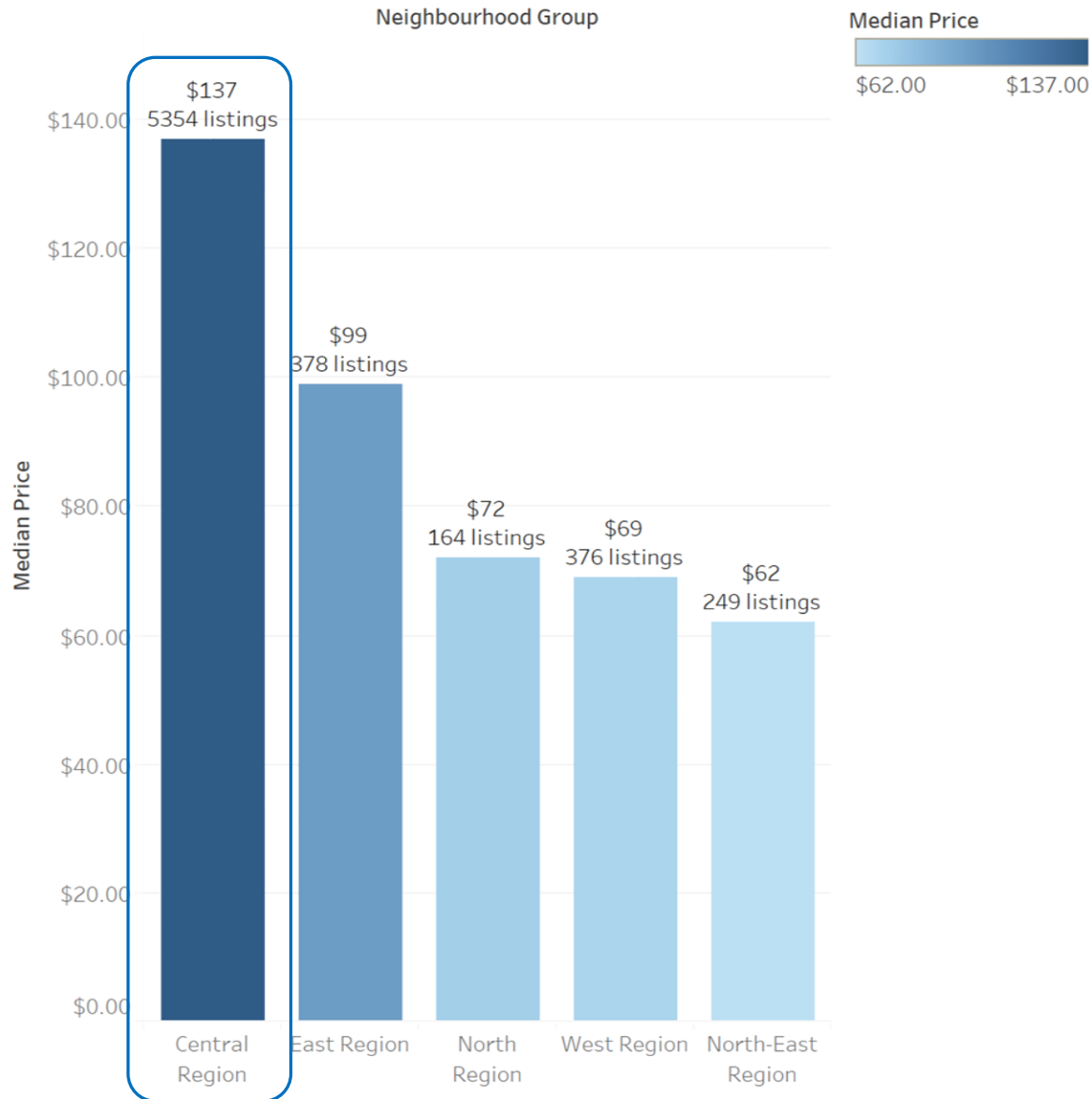
- What are the Possible Clusters?



# Key Findings I

Interesting Patterns of Listing

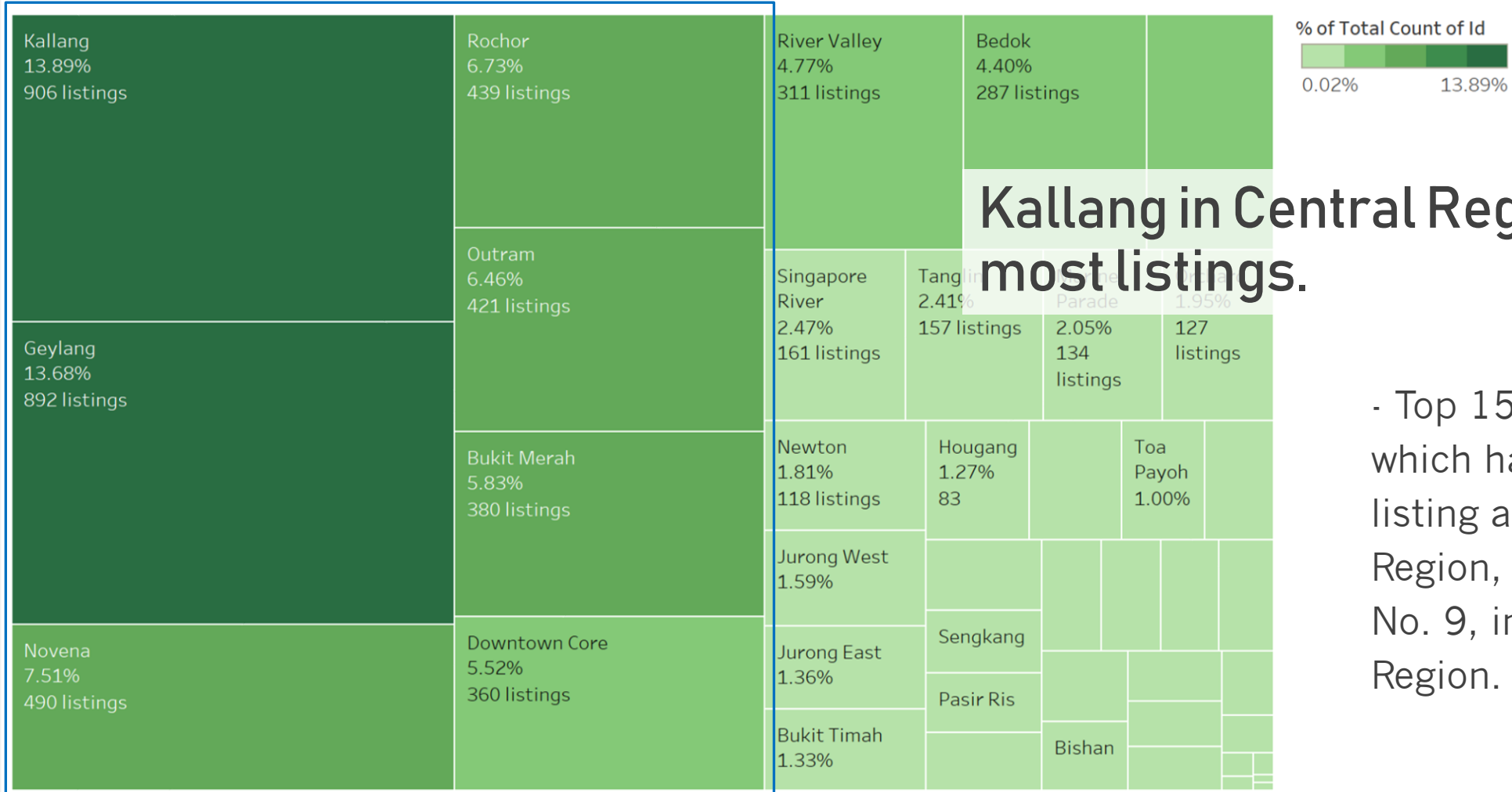
## Which Neighbourhood Group is Expensive?



**Apartments in the Central Region are the most expensive.**

- Followed by East Region, North Region, West Region and North-East Region.
- Top 1 exceeds top 2 by \$38, which is the biggest difference compared to that between other adjacent variables.

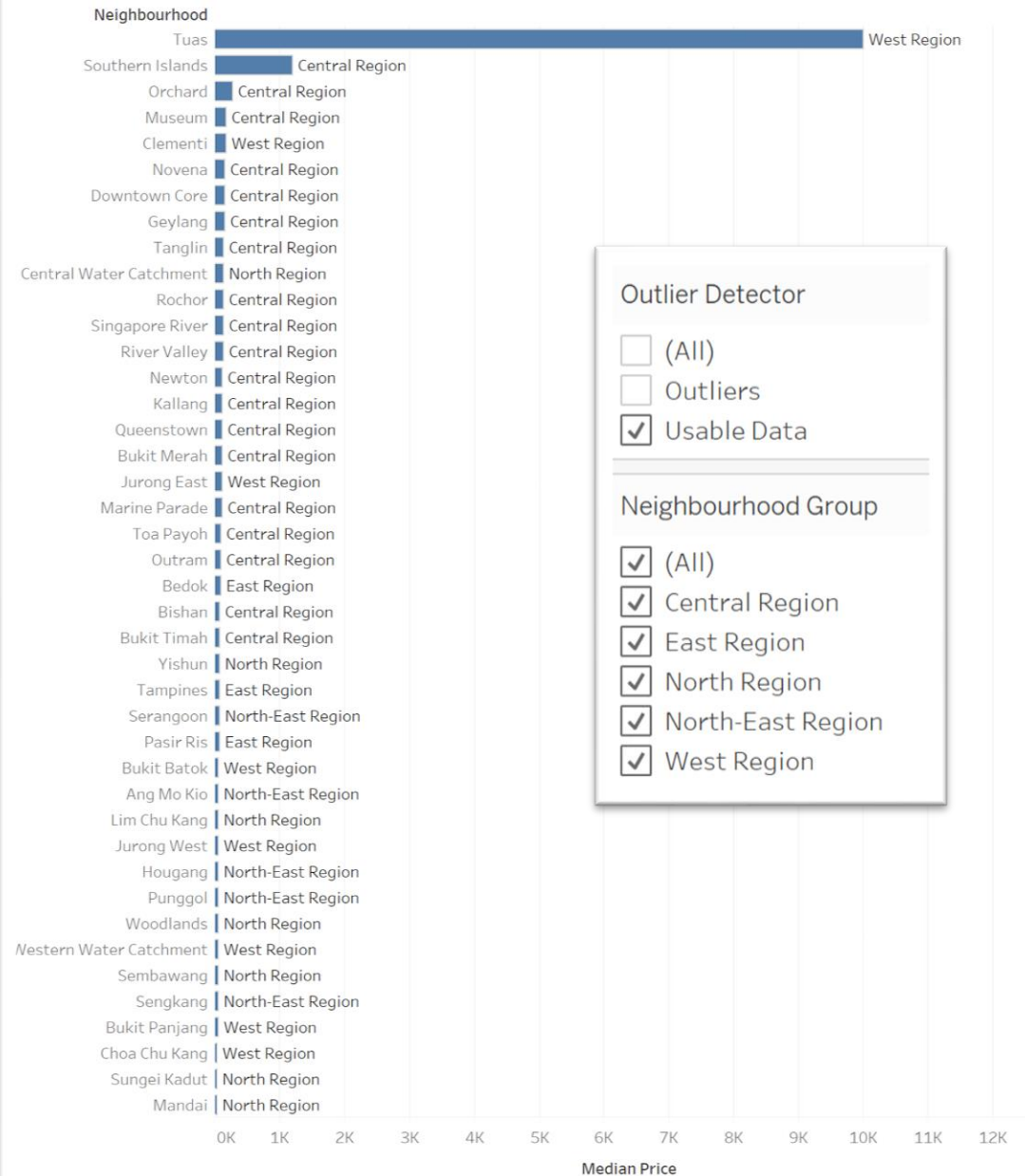
## Which Neighbourhood Has the Most Listings?



**Kallang in Central Region has the most listings.**

- Top 15 neighborhood which have the most listing are all in Central Region, except Bedok, No. 9, in the East Region.

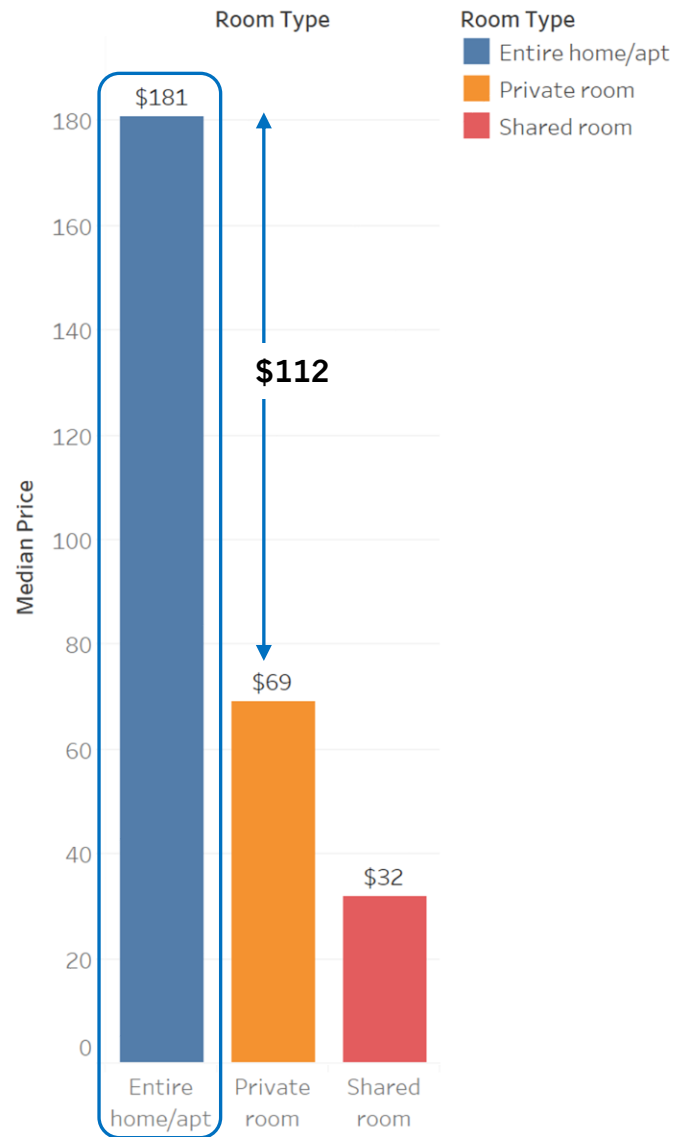
Which Neighbourhood is the Most Expensive in each Neighbourhood Group?



Neighborhood Tuas in West Region has apartments with the highest median price.

- Tuas in West Region.
- Southern Islands in Central Region.
- Bedok in East Region.
- Central Water Catchment in North Region.
- Serangoon in North-East Region.

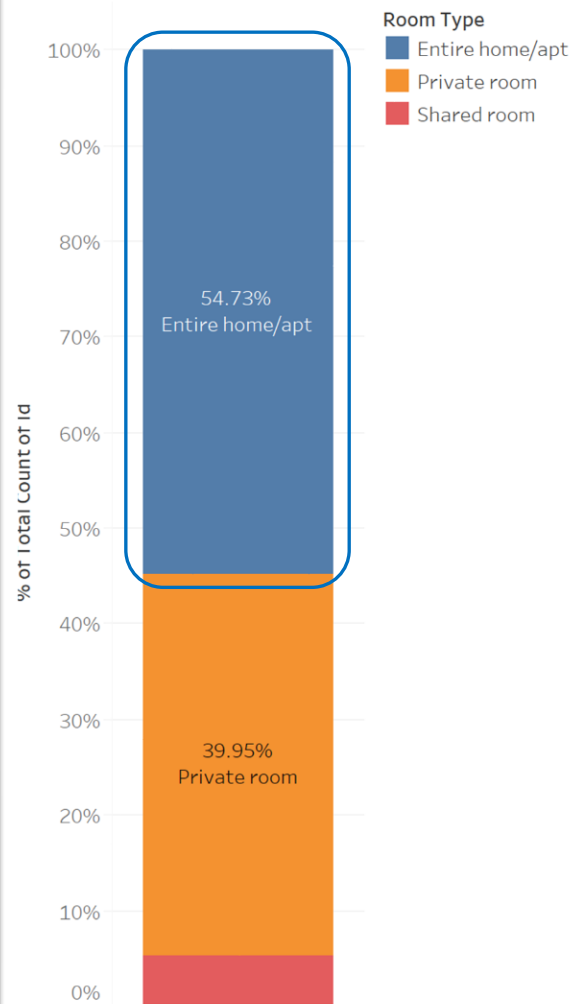
What is the Most  
Expensive Room Type?



**Entire apartment is the most expensive room type . Followed by private room and shared room.**

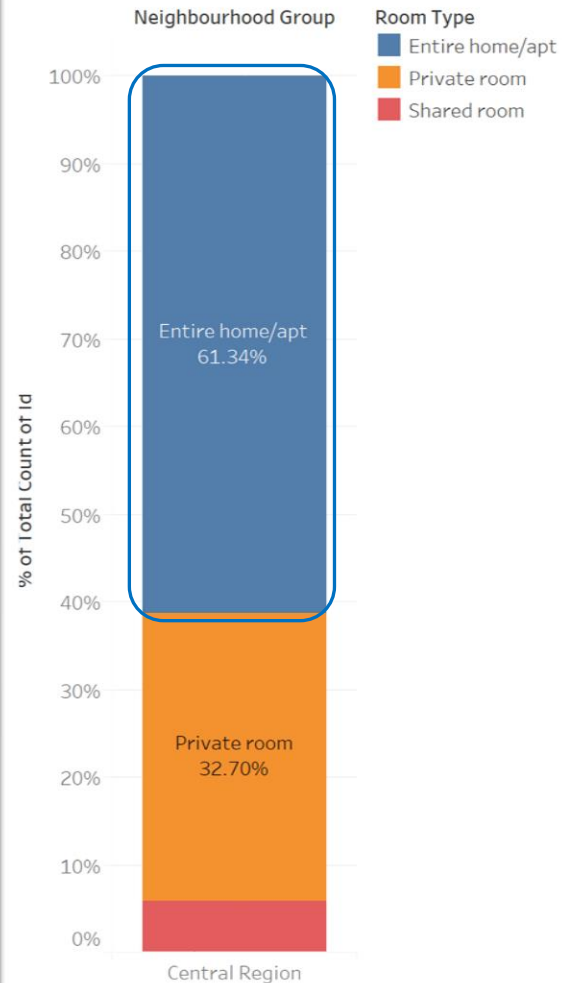
- Median price difference between the entire apt and private room reaches \$112.

## What is the Most Popular Room Type?



% of Total Count of Id. Color shows details about Room Type. The marks are labeled by % of Total Count of Id and Room Type. The data is filtered on Outlier Detector, which keeps Usable data.

## What is the Most Popular Room Type in Central Region?

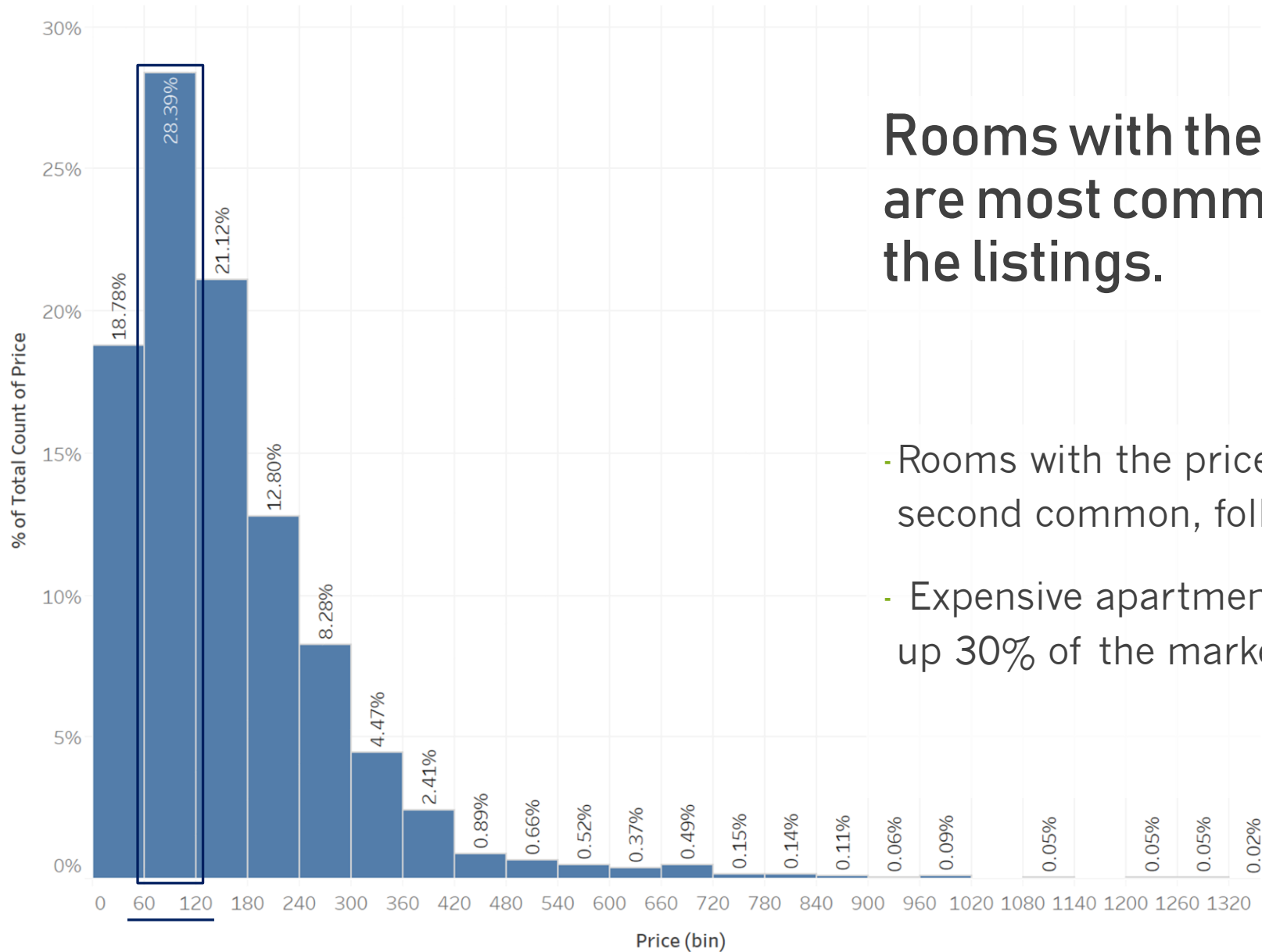


% of Total Count of Id for each Neighbourhood Group. Color shows details about Room Type. The marks are labeled by Room Type and % of Total Count of Id. The data is filtered

# Entire Home/Apartment is the most common room type.

- Hosts tend to rent out the whole apartments on Airbnb.
- Hosts in the Central Region are more likely to rent out their entire home on Airbnb compared to other groups.

What is the Price Range?



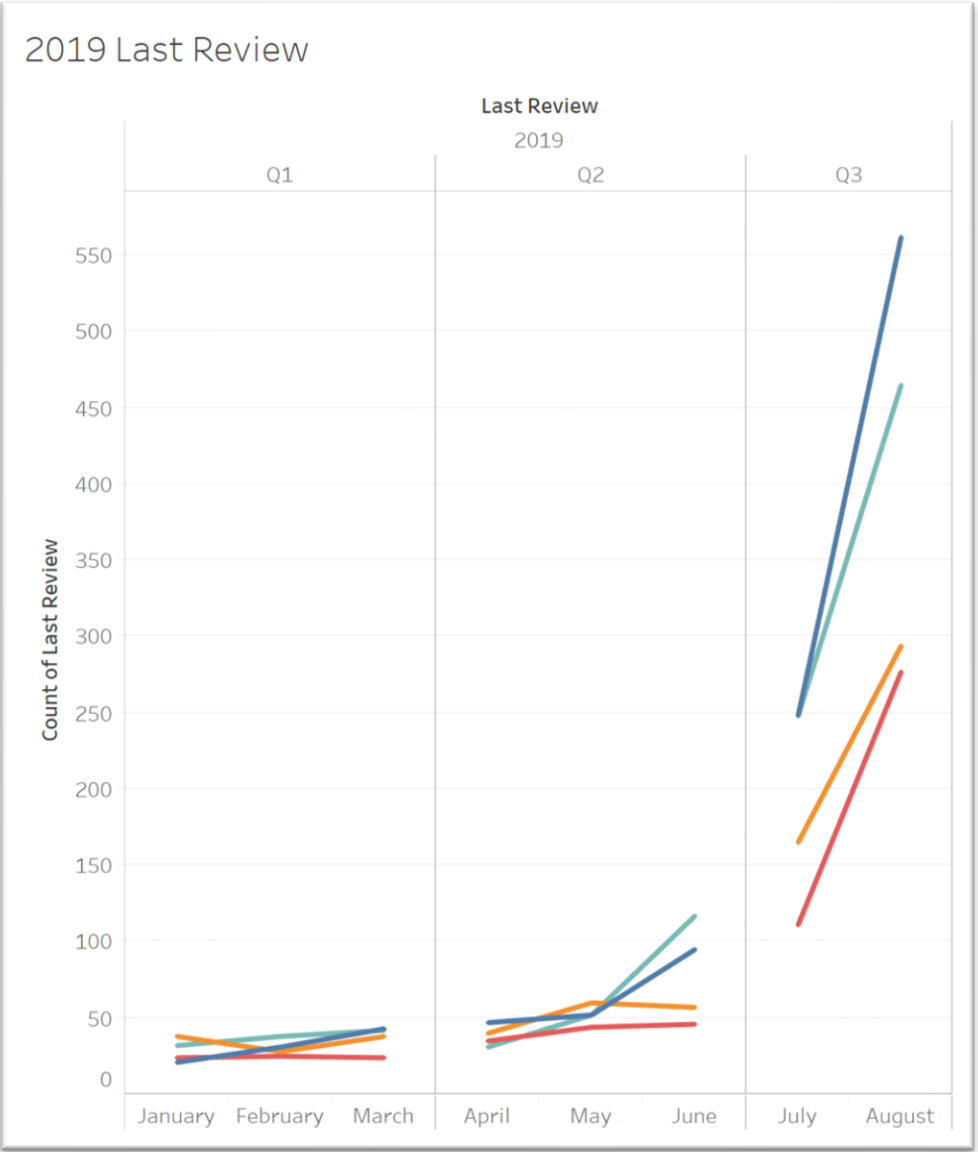
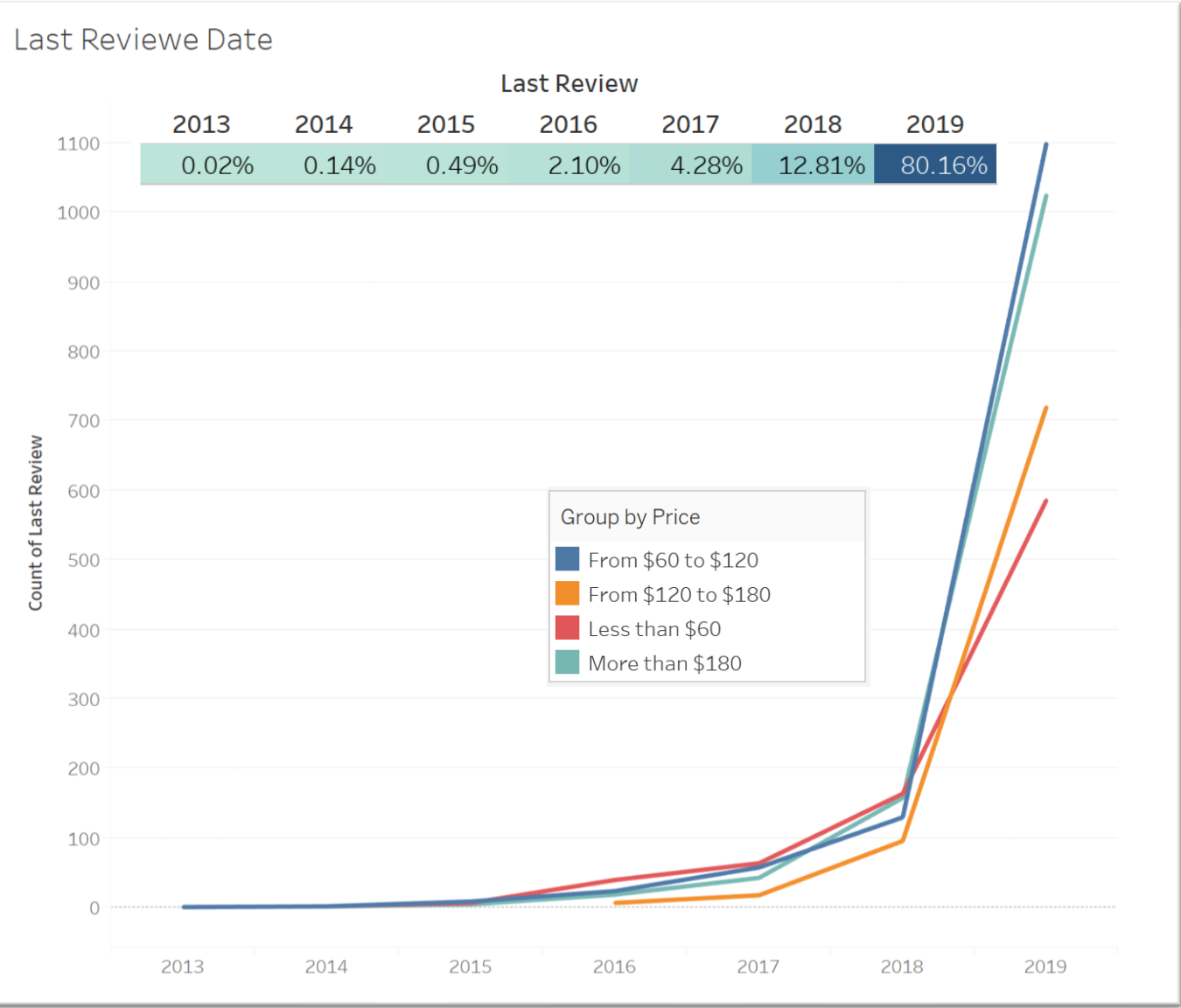
**Rooms with the price of \$60 to \$120 are most common . Nearly 30% of all the listings.**

- Rooms with the price of \$120 to \$180 are second common, followed by those below \$60.
- Expensive apartments (more than \$180) take up 30% of the market.



# Last Review Date Overview

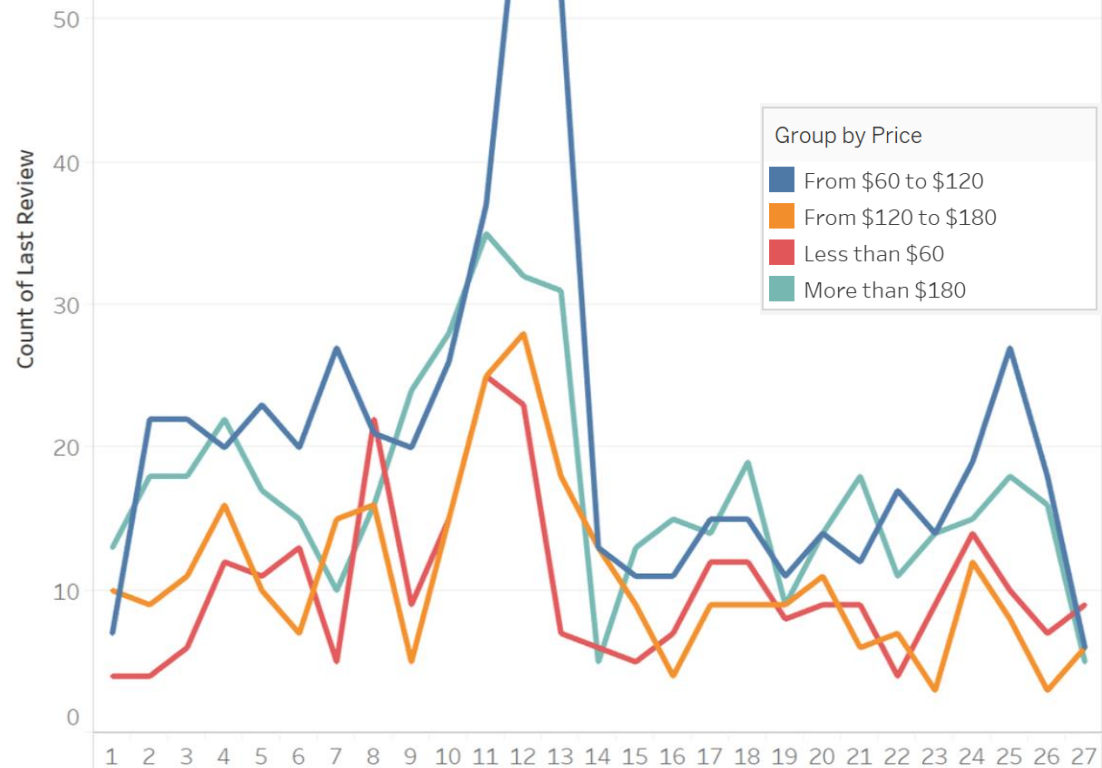
More than 80% of listings have recent reviews in 2019.



## August 2019 Last Review

Day of Last Review: 12  
Month of Last Review: August  
Group by Price: From \$60 to \$120  
Quarter of Last Review: Q3  
Year of Last Review: 2019  
Count of Last Review: 62

Last Review  
2019  
Q3  
August



## Reviews in Q3, 2019 surge

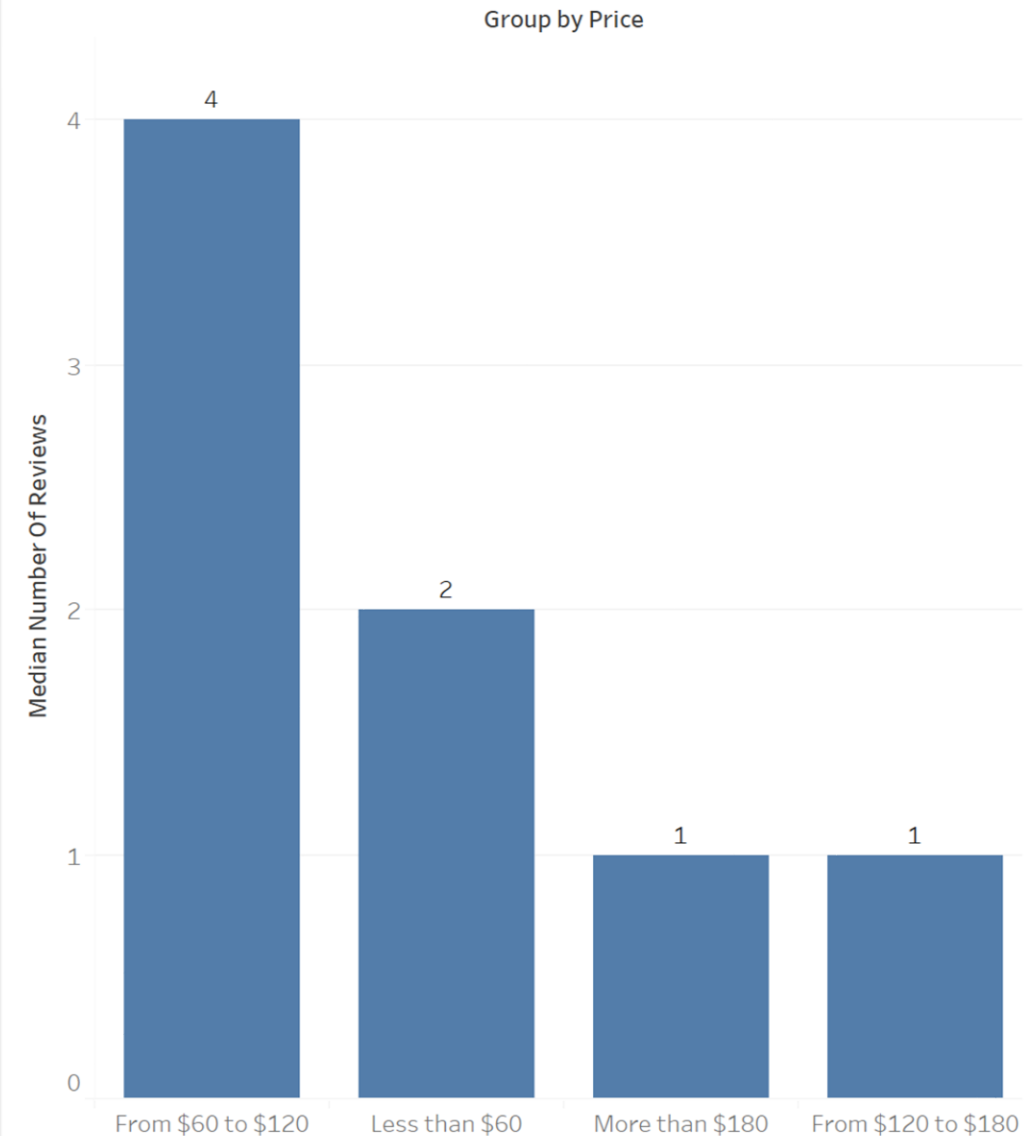
- Especially in August.
- On August 12<sup>th</sup>, 62 reviews appeared as the last review, which reaches the highest amount during the year.



# Key Findings II

What Factors May Correlate with Price?

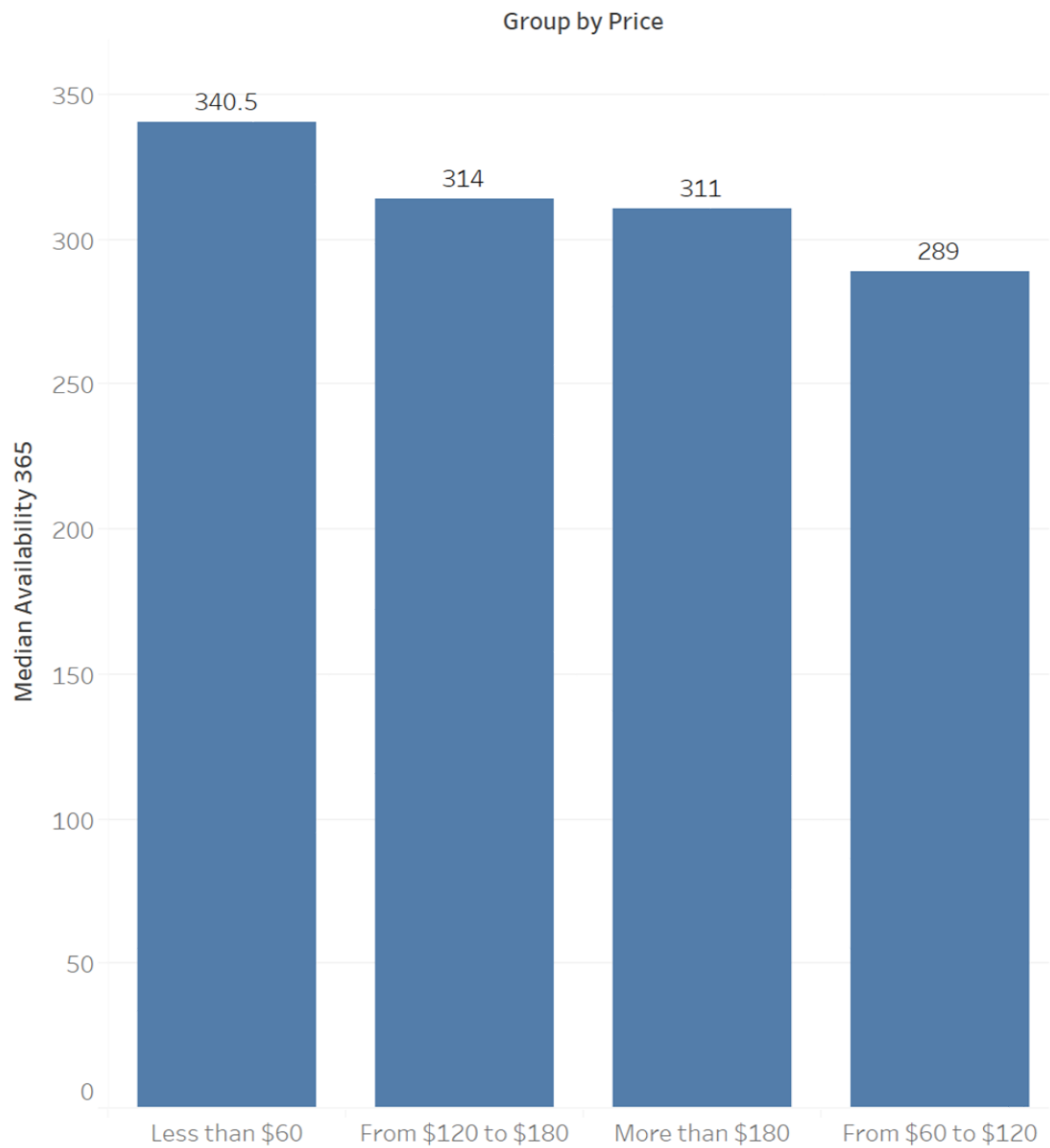
Do Apartments with More Reviews Tend to Have Higher Price Versus the other Groups?



**Rooms with lower-than-median price tend to have more reviews.**

- On the opposite, rooms with the higher-than-median price tend to have less reviews.

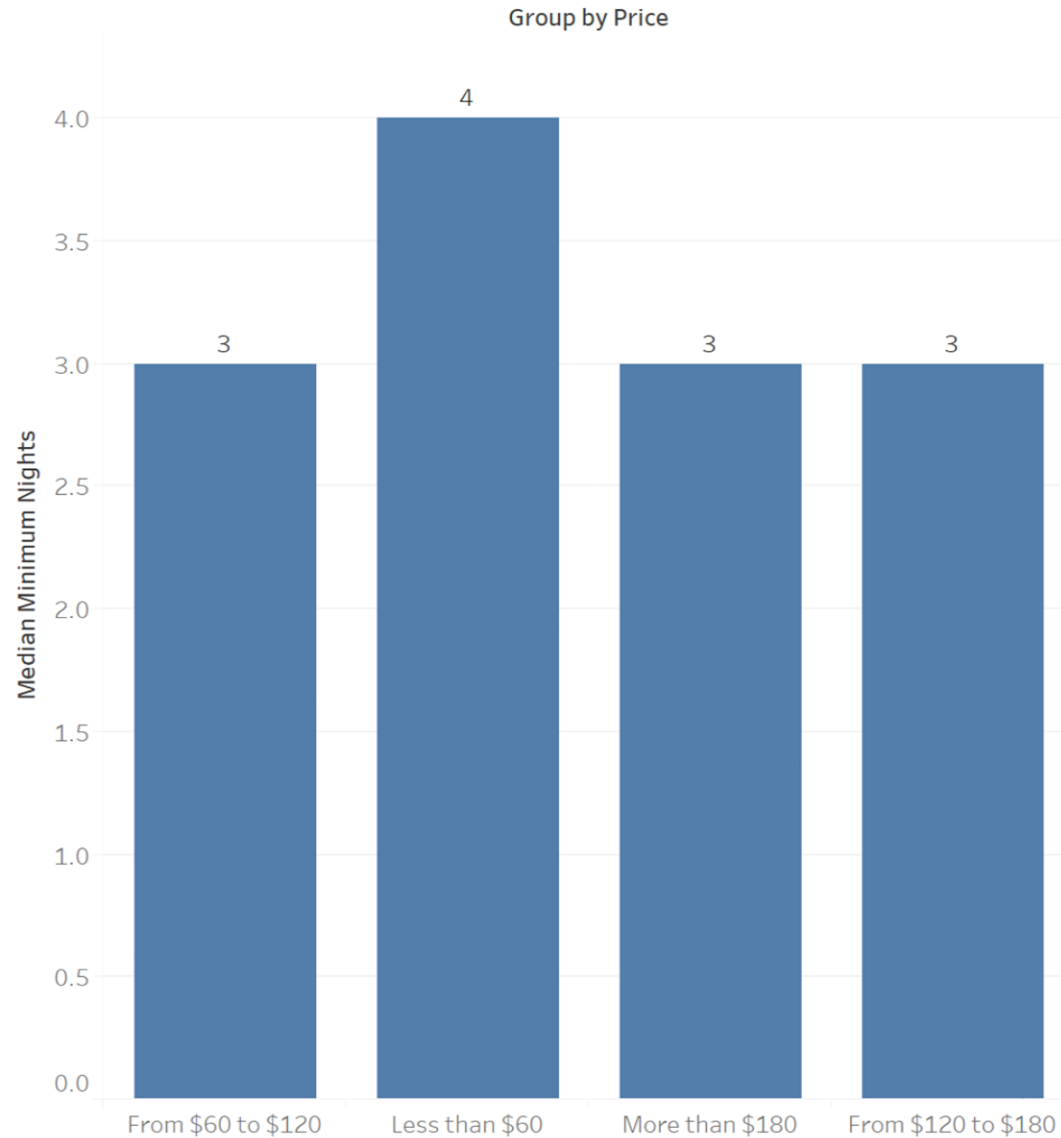
Do Apartments with More Availabilities Tend to Have Higher Price Versus the other Groups?



Rooms with the price less than \$60 tend to have the most availabilities.

- The \$60 to \$120 group has the least availability.

Do Apartments with Less Min Nights Tend to Have Higher Price Versus the other Groups?



**Cheap rooms tend to have more required minimum nights.**

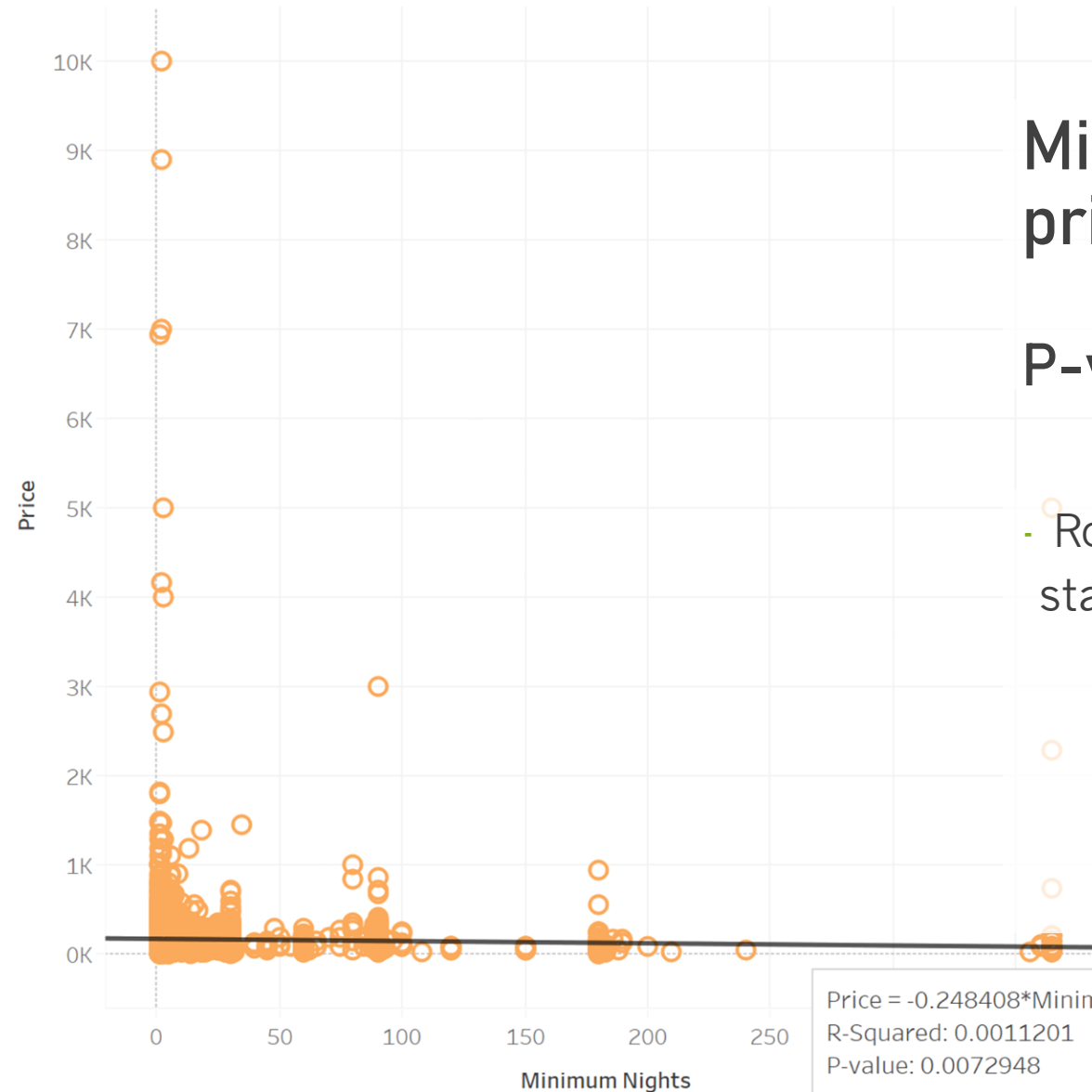
- The difference is not very obvious in other groups.



# Key Findings III

What are the Key Factors Impacting the Price?

## #Min Night & Price



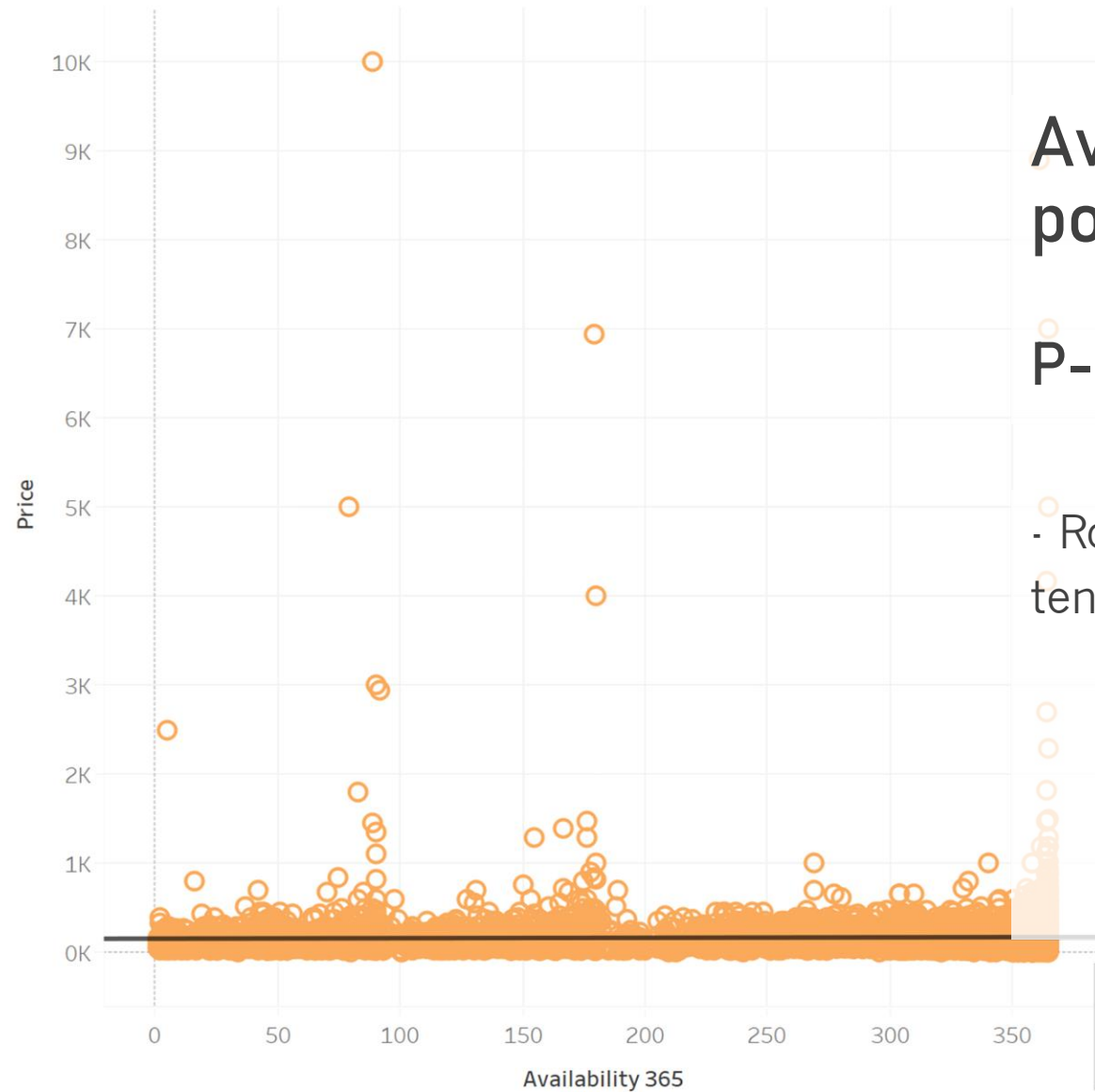
Minimum night requirement and price have negative correlation.

P-value: 0.007

Rooms with more minimum nights needed to stay tend to be cheaper, vice versa.



## #Availability & Price



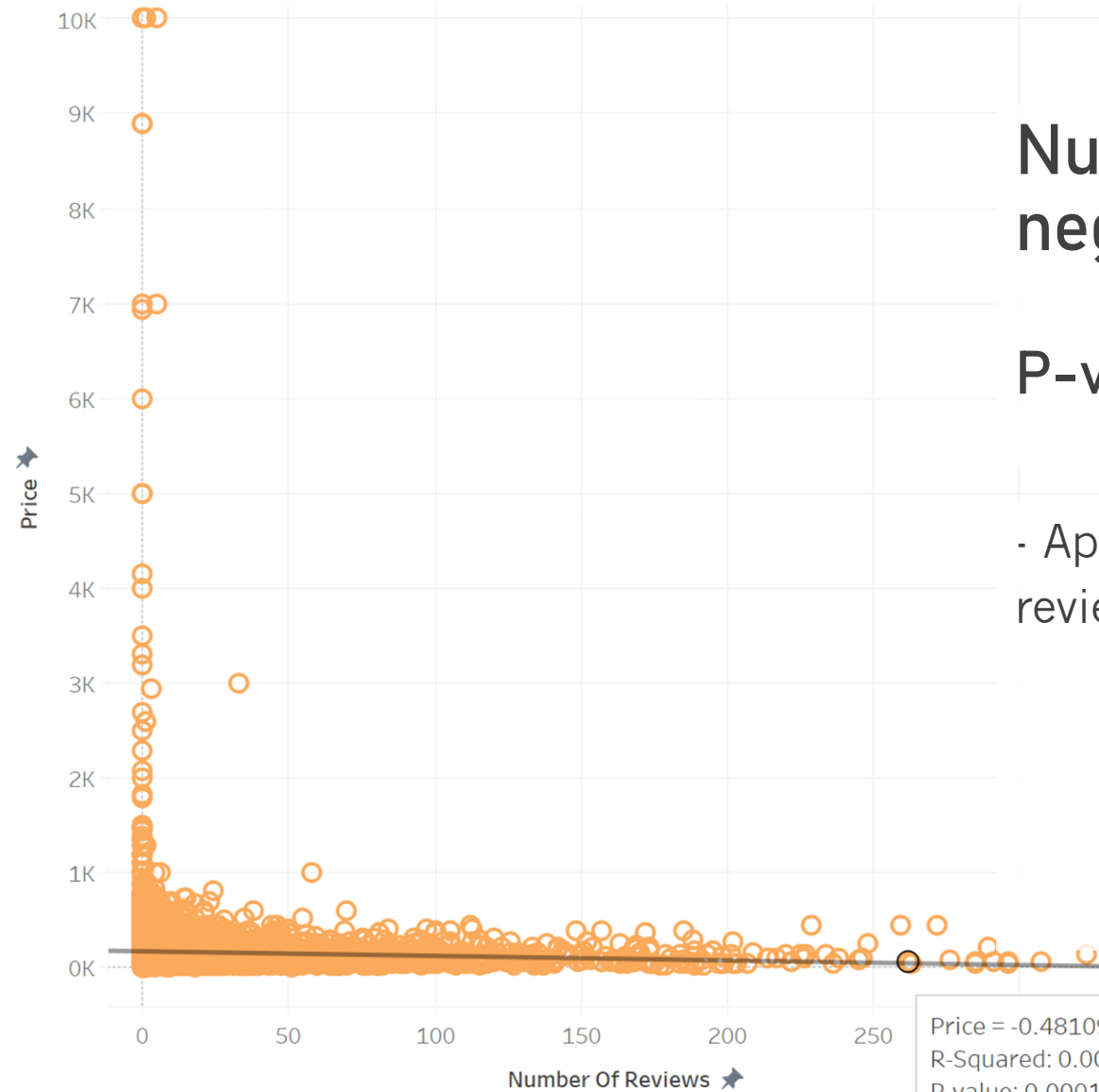
Available days and price have positive correlation.

P-value: 0.113

- Rooms with more availabilities around the year tend to be more expensive, vice versa.

Price = 0.0461769\*Availability 365 + 159.446  
R-Squared: 0.000391  
P-value: 0.112987

## #Reviews & Price



Number of reviews and price have negative correlation.

P-value: 0.0002

- Apartments with lower price tend to have more reviews, vice versa.

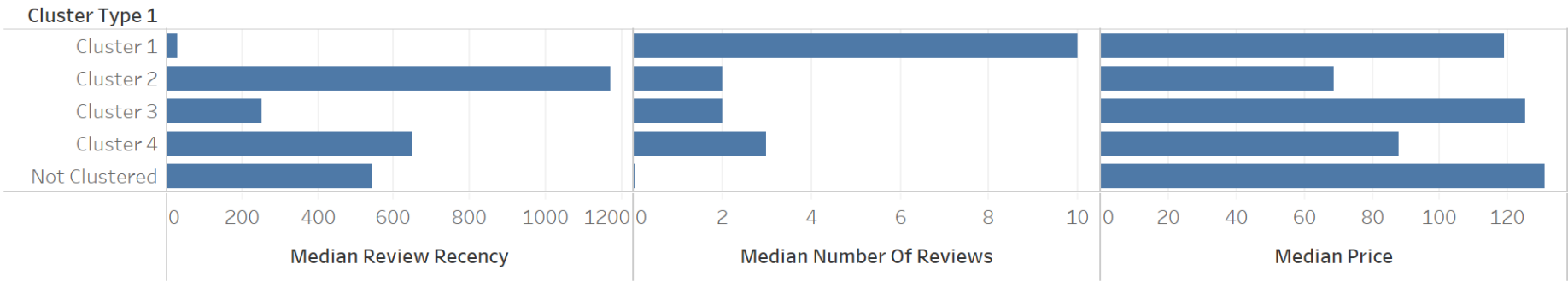
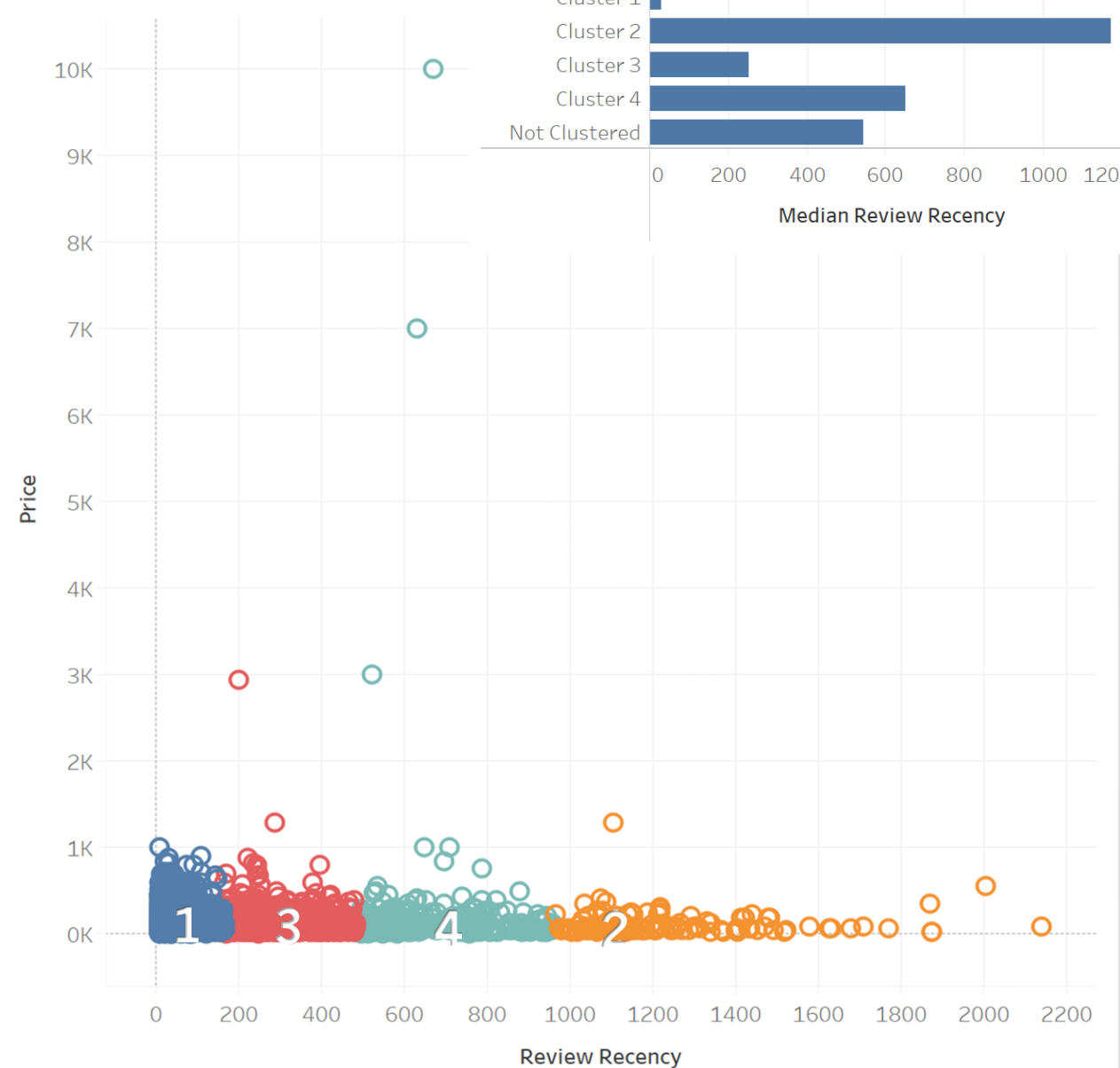
Price = -0.481097\*Number Of Reviews + 175.495  
R-Squared: 0.0017651  
P-value: 0.0001863



# Key Findings IV

Listing Segmentation

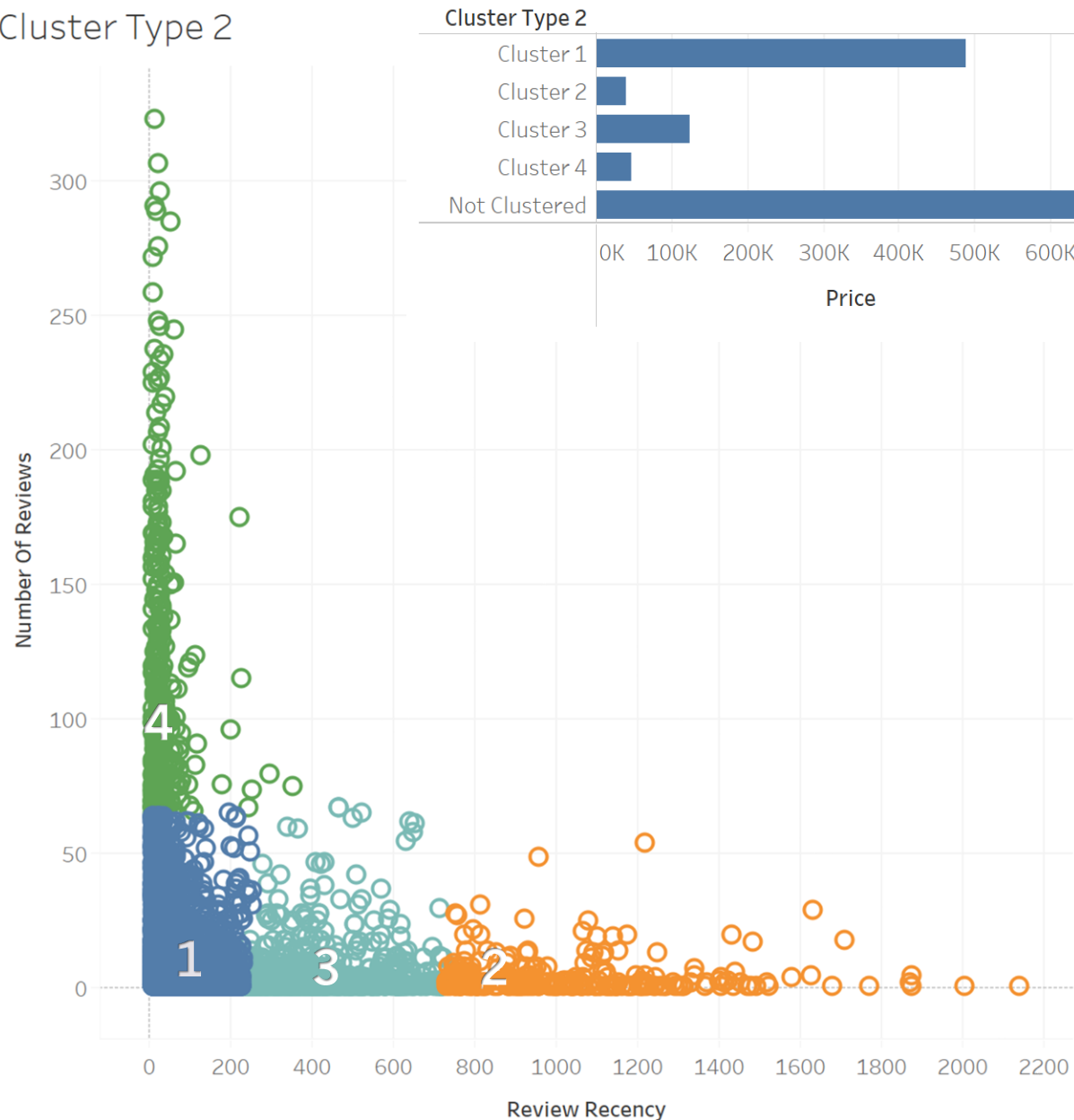
Cluster Type I



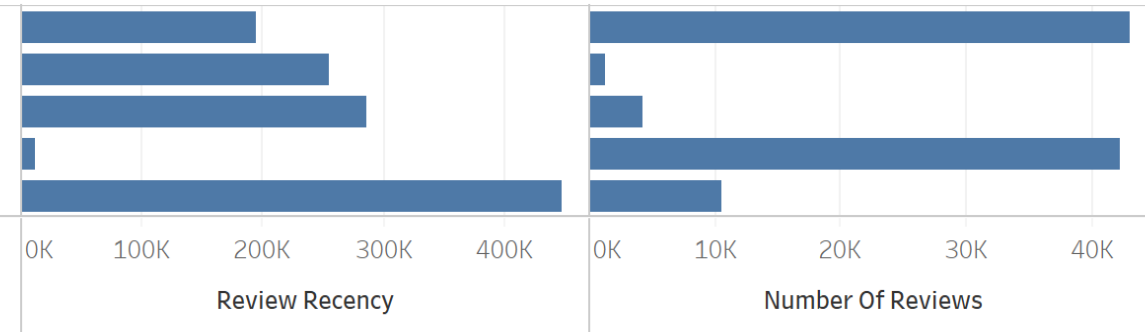
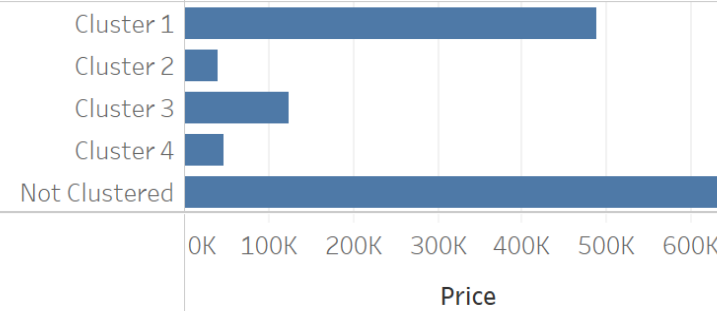
Cluster Type 1	Median Number Of Reviews	Median Price	Median Review Recency
Cluster 1	10	119	30
Cluster 2	2	69	1,173
Cluster 3	2	125	254
Cluster 4	3	88	649
Not Clustered	0	131	541

- Rooms with lower price tend to have weak recency performance.
- Rooms with more number of reviews tend to have better recency performance.

## Cluster Type 2



### Cluster Type 2



Cluster Type 2	Median Number Of Reviews	Median Price	Median Review Recency
Cluster 1	7.0	125.0	36.0
Cluster 2	2.0	69.0	975.5
Cluster 3	2.0	118.0	382.0
Cluster 4	99.0	96.0	21.0
Not Clustered	0.0	131.0	632.0

- Review recency performance rank
  - Cluster 4 > 1 > 2 > 3
- Rooms with more number of reviews tend to have better recency performance.

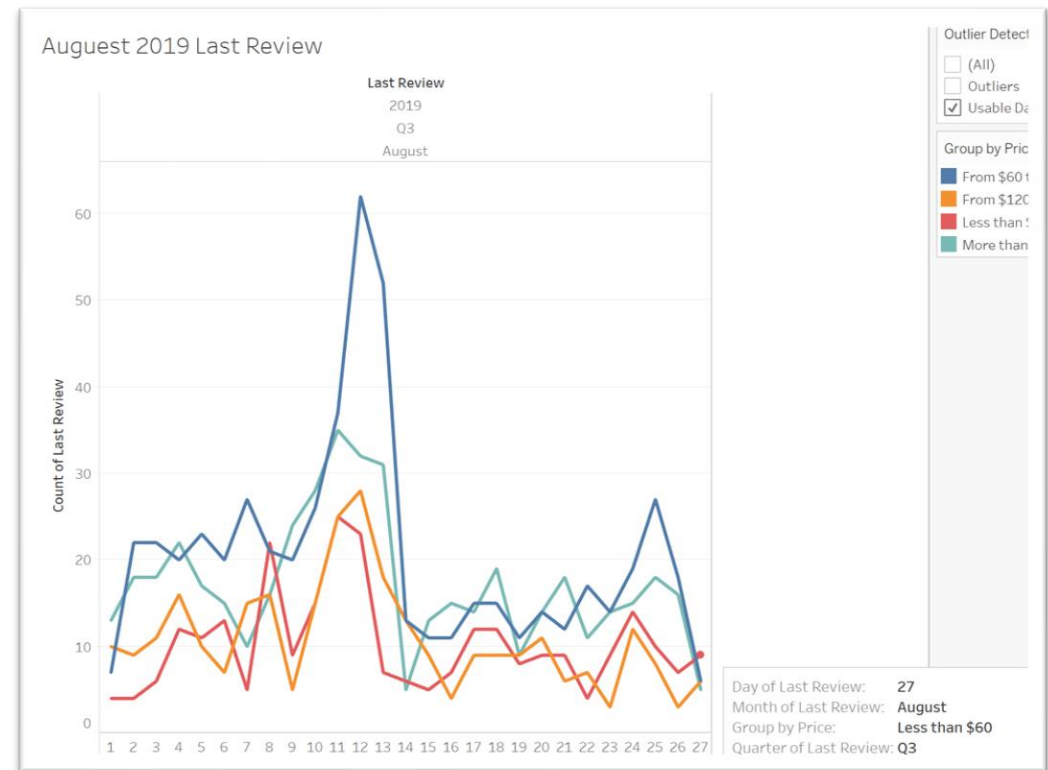
# Methodology I

create review recency

- Figure out the most recent review date
  - Aug. 27<sup>th</sup>, 2019
- Create new calculated field
  - Sep. 1<sup>st</sup>, 2019
- Create new calculated field
  - Review Recency (days before Sep. 1<sup>st</sup>)

Review Recency

```
DATEDIFF('day', [Last Review], [Sep. 1st] )
```

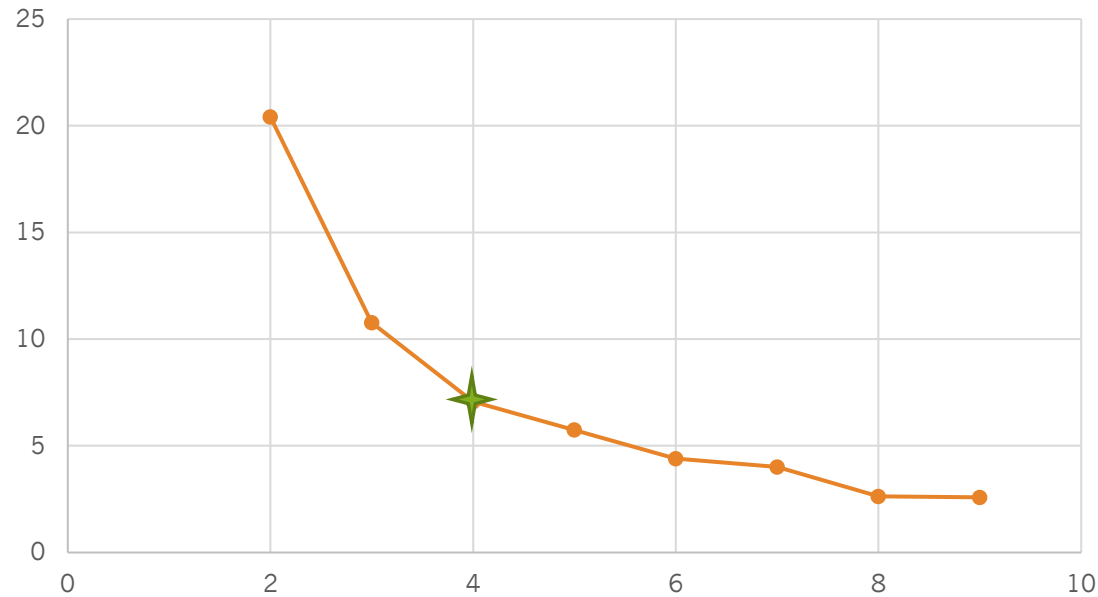


# Methodology II

decide elbow point

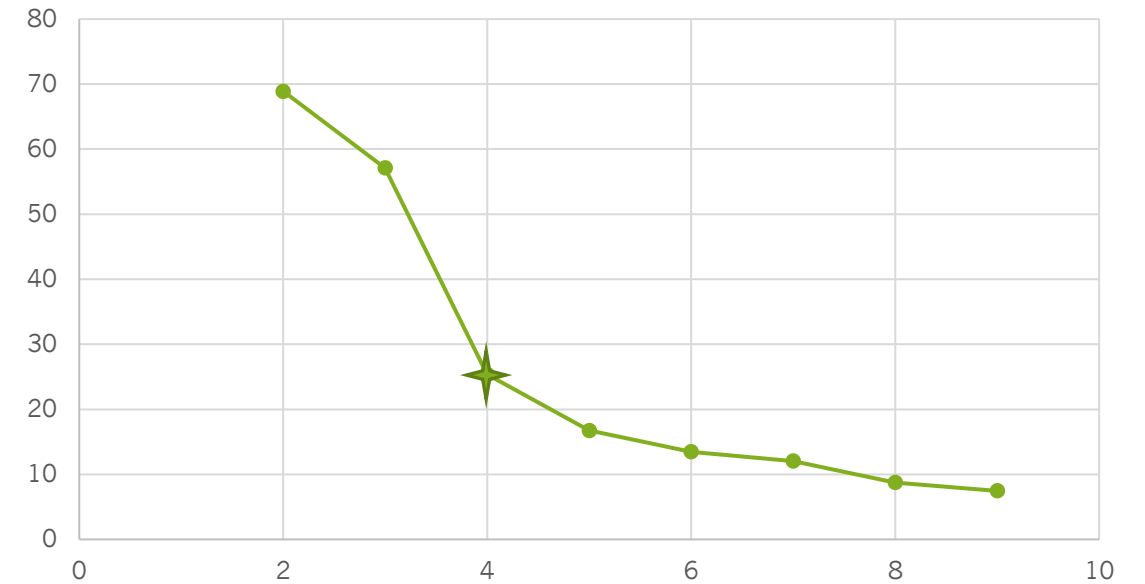
---

Clusters & Within-Group Sum of Squares



Cluster Type 1

Within-Group Sum of Squares



Cluster Type 2

# Summary

---

- Neighborhood, neighborhood group and room type impact the price of the listing.
- Cheap rooms tend to have more required minimum nights.
- Rooms with more availabilities around the year tend to be more expensive, vice versa.
- Rooms with lower price tend to have more reviews, vice versa.
- Rooms with more number of reviews tend to have better recency performance.





# Thanks for Your Attention!

Manyu Jiang