

Airbnb: correlation between distance and price

VADIM ACOSTA

At a glance

- ▶ Airbnb Data from Kaggle
- ▶ **48895 records**, trimmed to **31014 records** to exclude nulls
- ▶ 16 variables including host id, host name, number of listings, number of reviews, GPS coordinates, price and others ...
- ▶ Tools: Tableau and Excel



FOCUS

- ▶ Boroughs
- ▶ Latitude and Longitude Data
- ▶ **Prices**

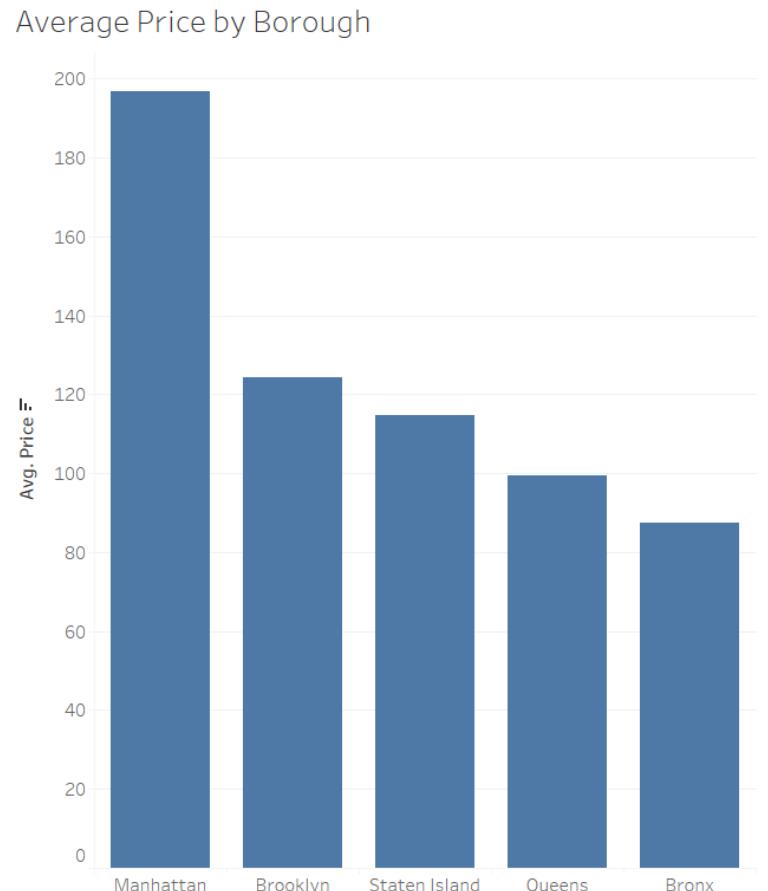
Mean	152.7207
Standard Error	1.08607
Median	106
Mode	100
Standard Deviation	240.1542
Sample Variance	57674.03
Kurtosis	585.6729
Skewness	19.11894
Range	10000
Minimum	0
Maximum	10000
Sum	7467278
Count	48895
Confidence Level(95.0%)	2.128711

Initial Insight

Null Hypothesis: There are no significant differences in price between boroughs

Alternative Hypothesis: There are significant differences in price between boroughs

Test Shows: with an alpha of 0.05 and a p-value of approximately 0.00 the null hypothesis is rejected and the **alternative hypothesis is confirmed**



Insight: Different places have different prices

- ▶ Can we express this insight as a mathematical function or regularity?
- ▶ Can we apply this insight in other markets?

?

New Alternative Hypotheses

- 1) Increased distance from Times Square* **correlates** with **cheaper** accommodations
- 2) Increased distance from Times Square* **correlates** with **fewer** accommodations

underlying insight: price can be modeled as a function of distance from an **arbitrary or preselected*** location

distances between points were not part of the dataset and had to be derived

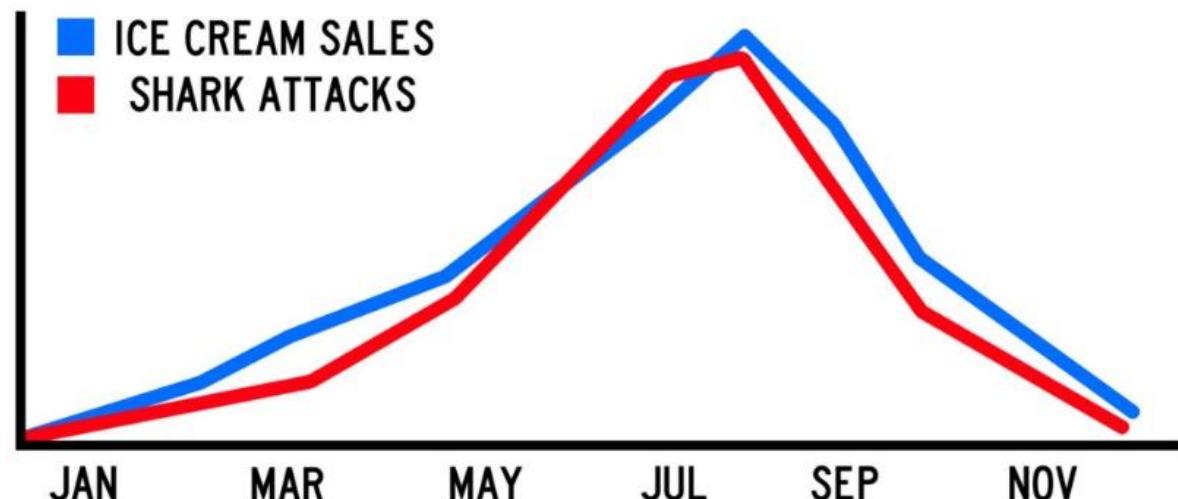


$$\begin{aligned} &= 6371 * \text{ACOS}(\text{COS}(\text{RADIAN}(90 - 40.758)) * \text{COS}(\text{RADIAN}(90 - D2)) \\ &+ \text{SIN}(\text{RADIAN}(90 - 40.758)) * \text{SIN}(\text{RADIAN}(90 - D2)) * \text{COS}(\text{RADIAN}((-73.9855) - E2))) / 1.609 \end{aligned}$$

Correlation

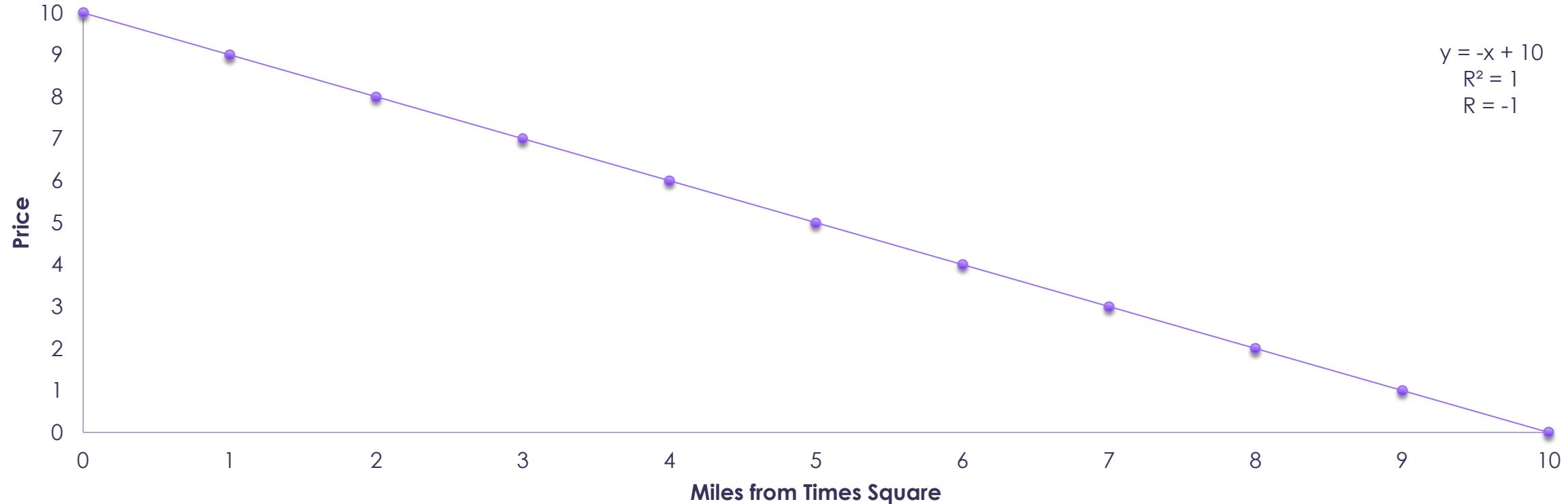
- ▶ R - has values between **-1** and **1**, measures the linear relationship between two quantitative variables with respect to direction and strength
- ▶ R^2 – has values between **0** and **1**, a measure of how close each data point fits to the regression line, how well the regression line predicts values

CORRELATION IS NOT CAUSATION!

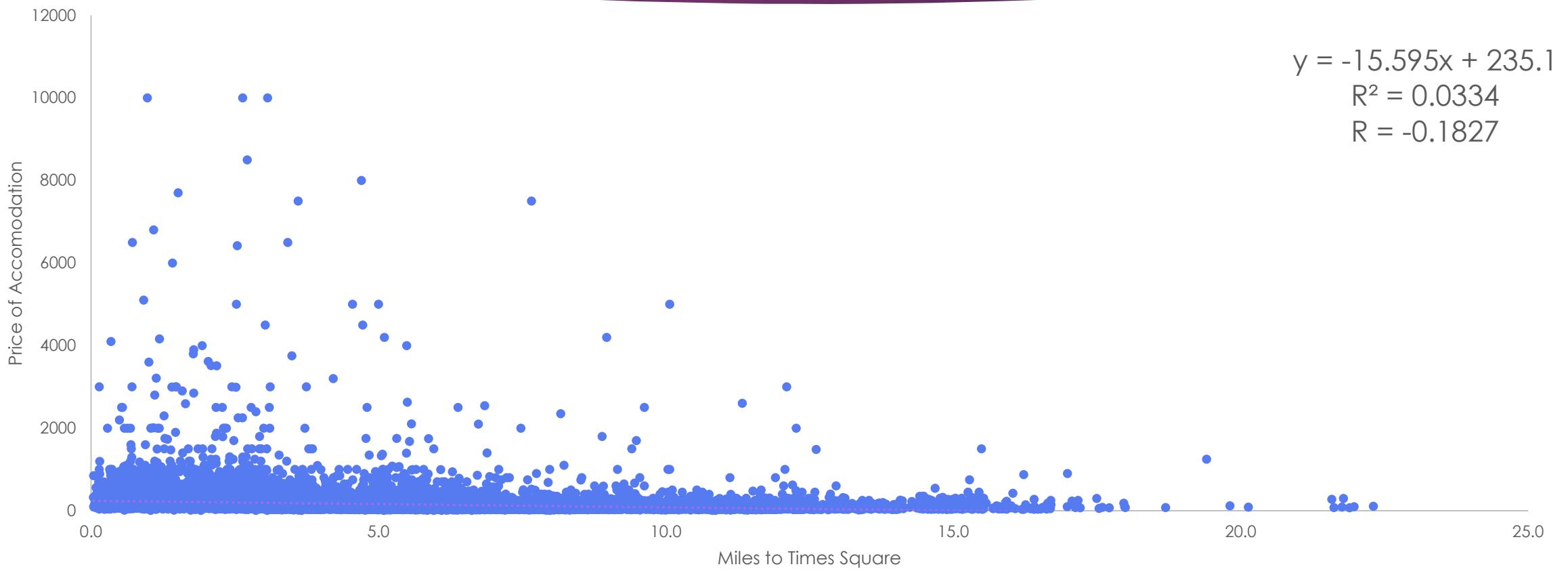


Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

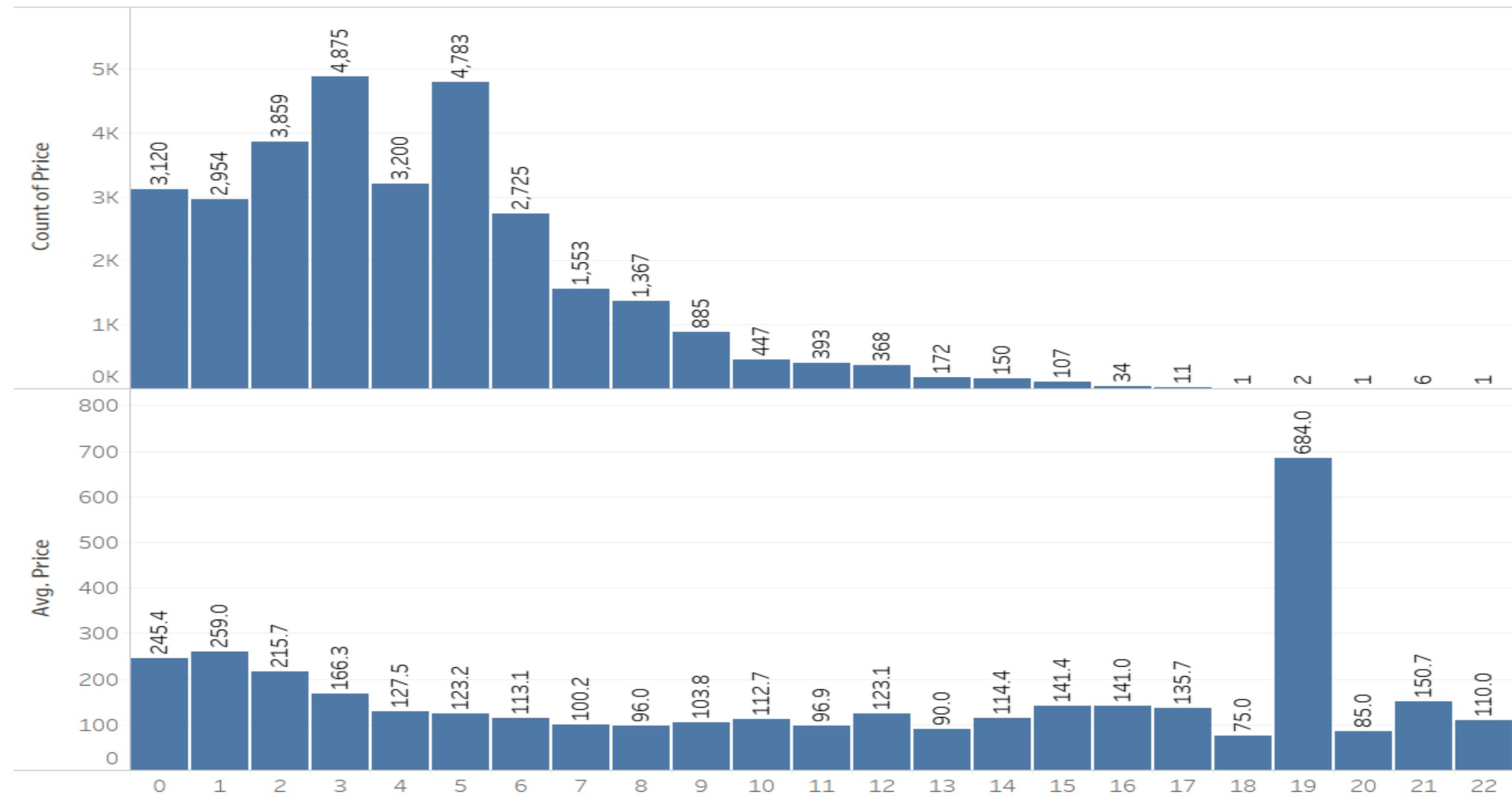
Ideal Inverse Correlation



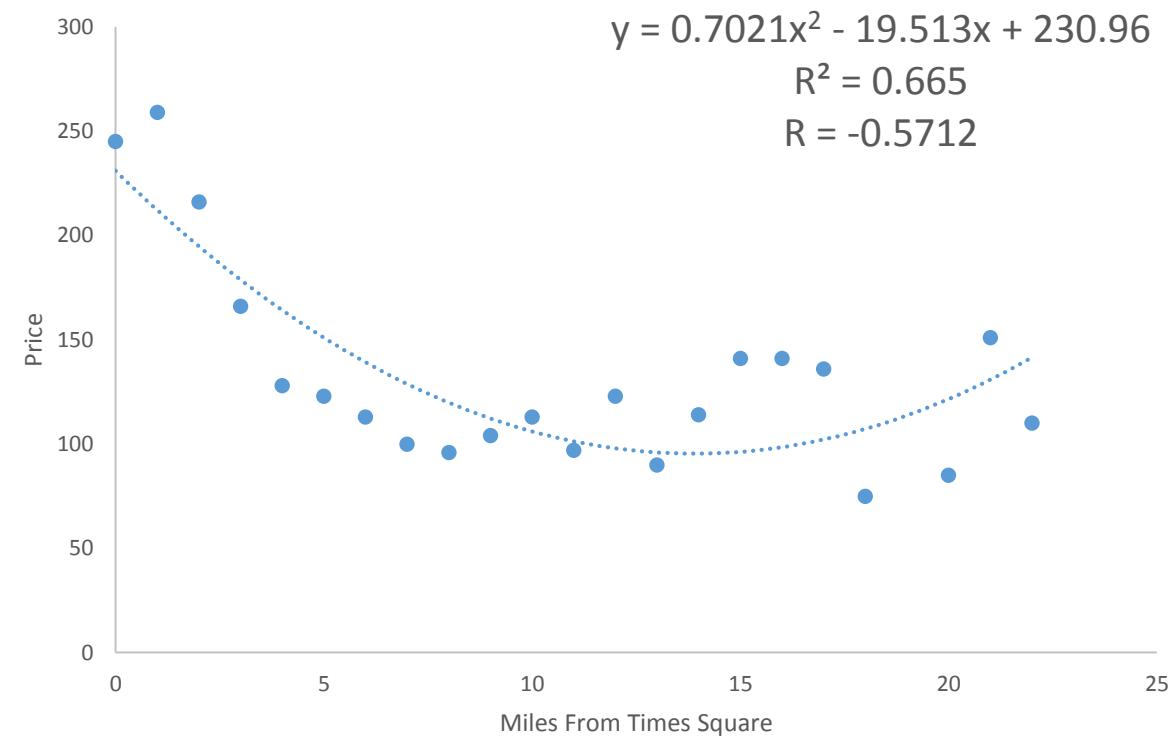
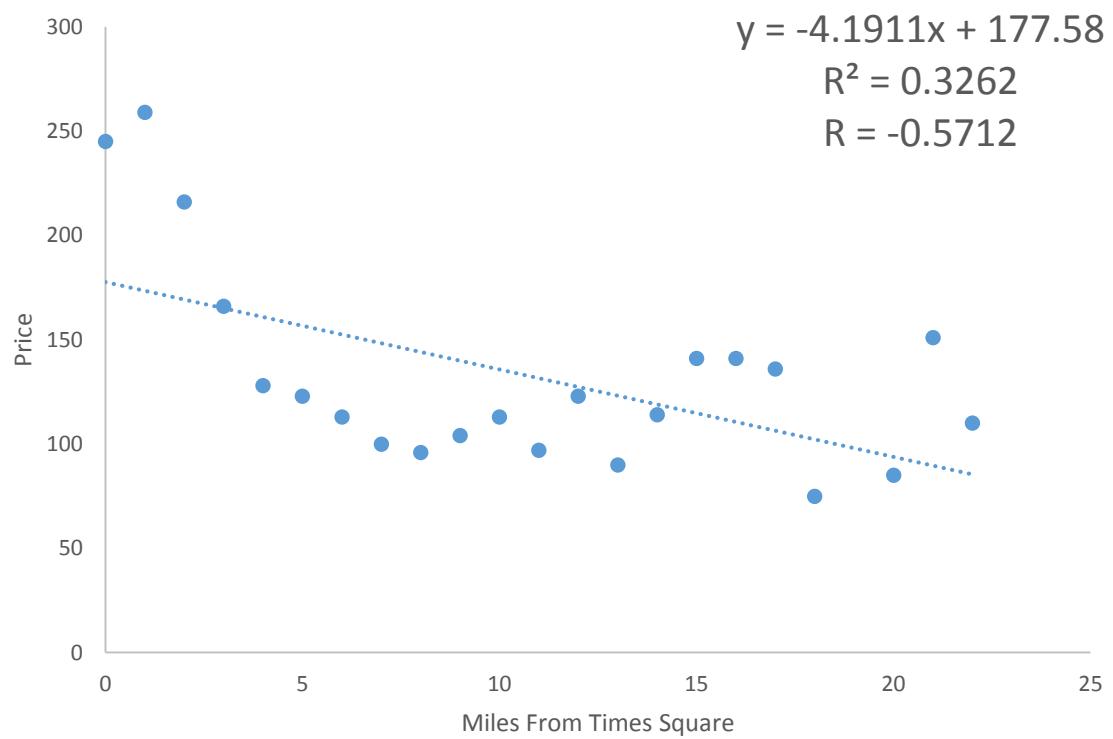
Price per mile



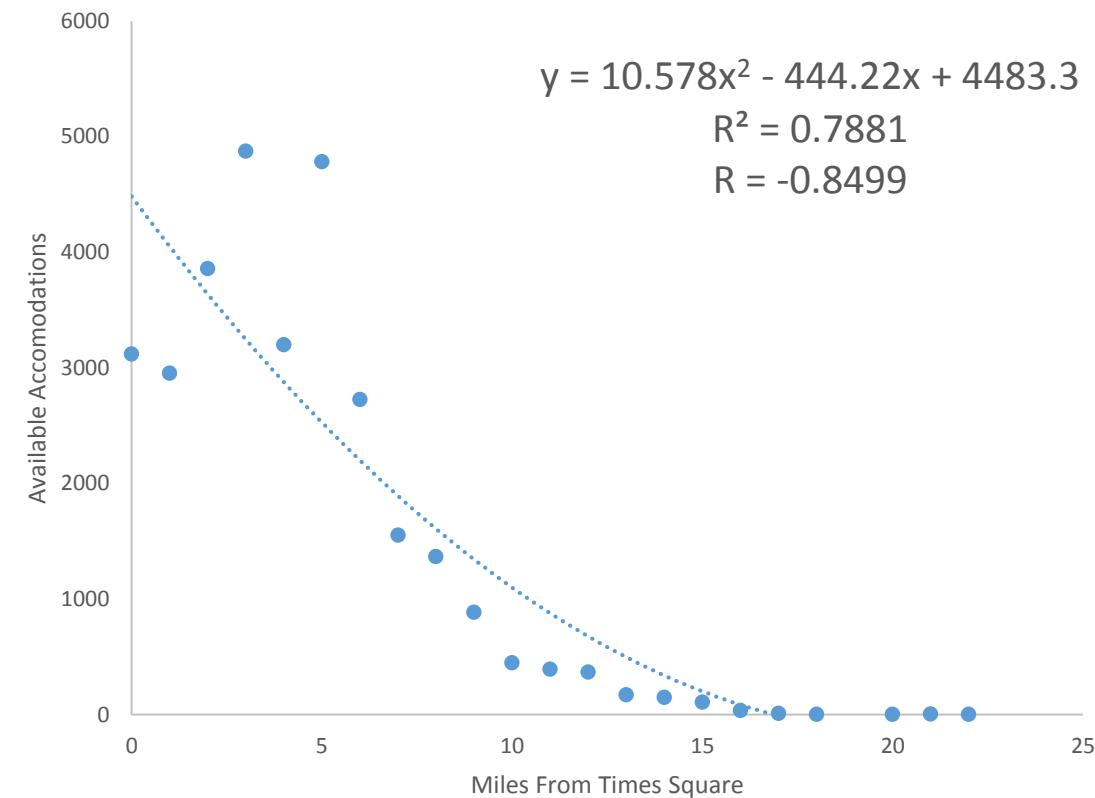
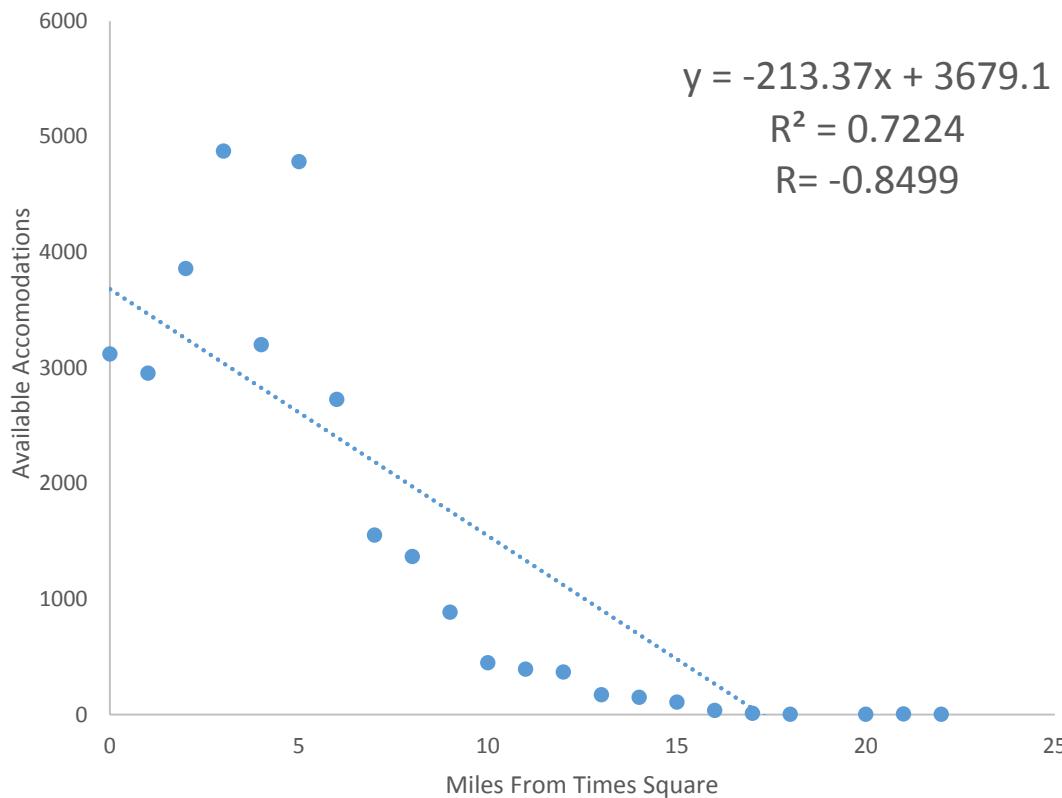
All Accomodations / Miles to Times Square



Average Price



Average Number of Accommodations



Conclusions and future research

- ▶ Prices can be modeled as a function of geography
- ▶ Geographical analysis doesn't require knowledge of underlying causes of price distribution
- ▶ Can be used to pinpoint influential locations that correlate with price

Thank you!

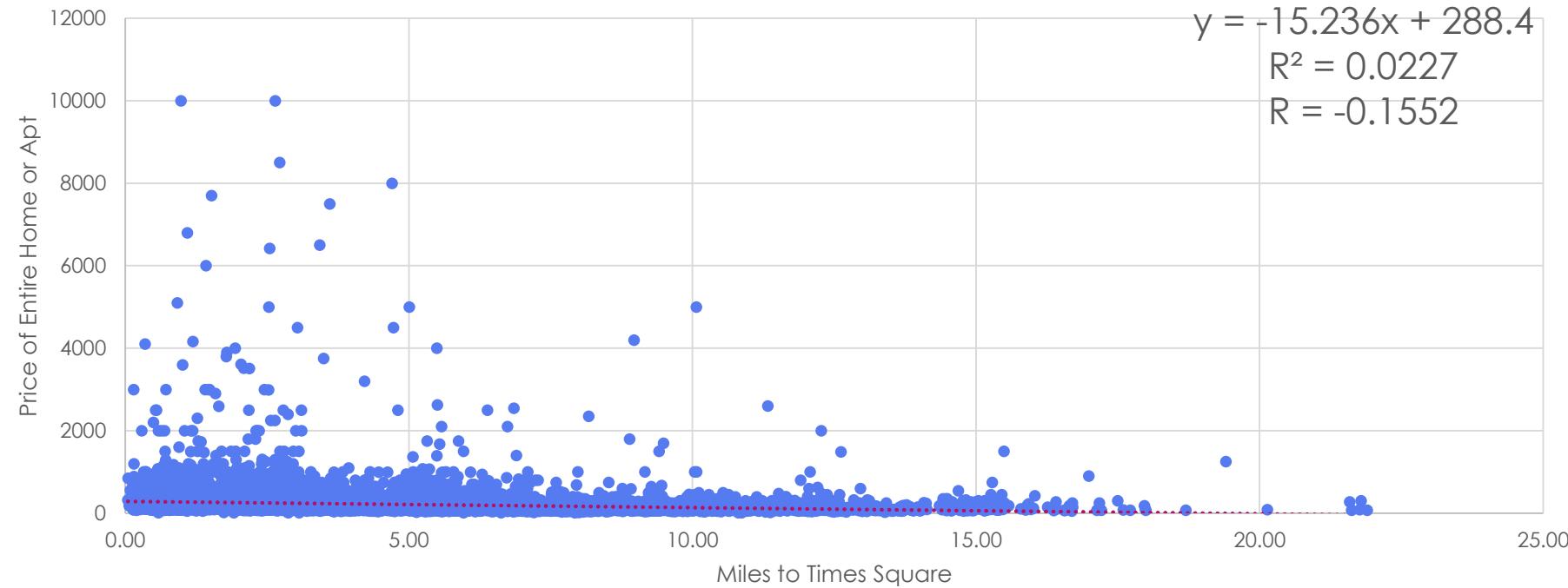


**NYC
Tech Talent
Pipeline**

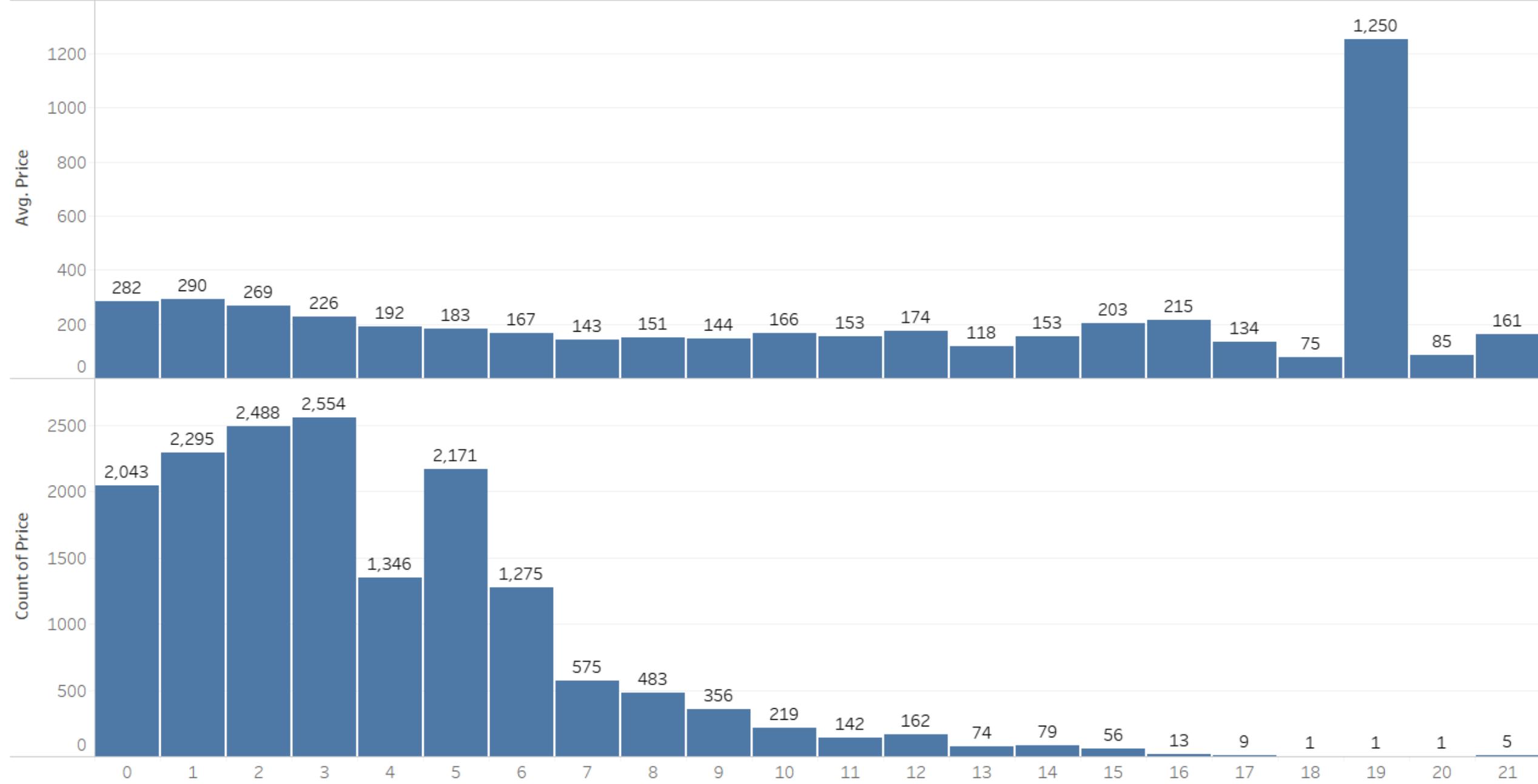
galvanize

Extra Slides

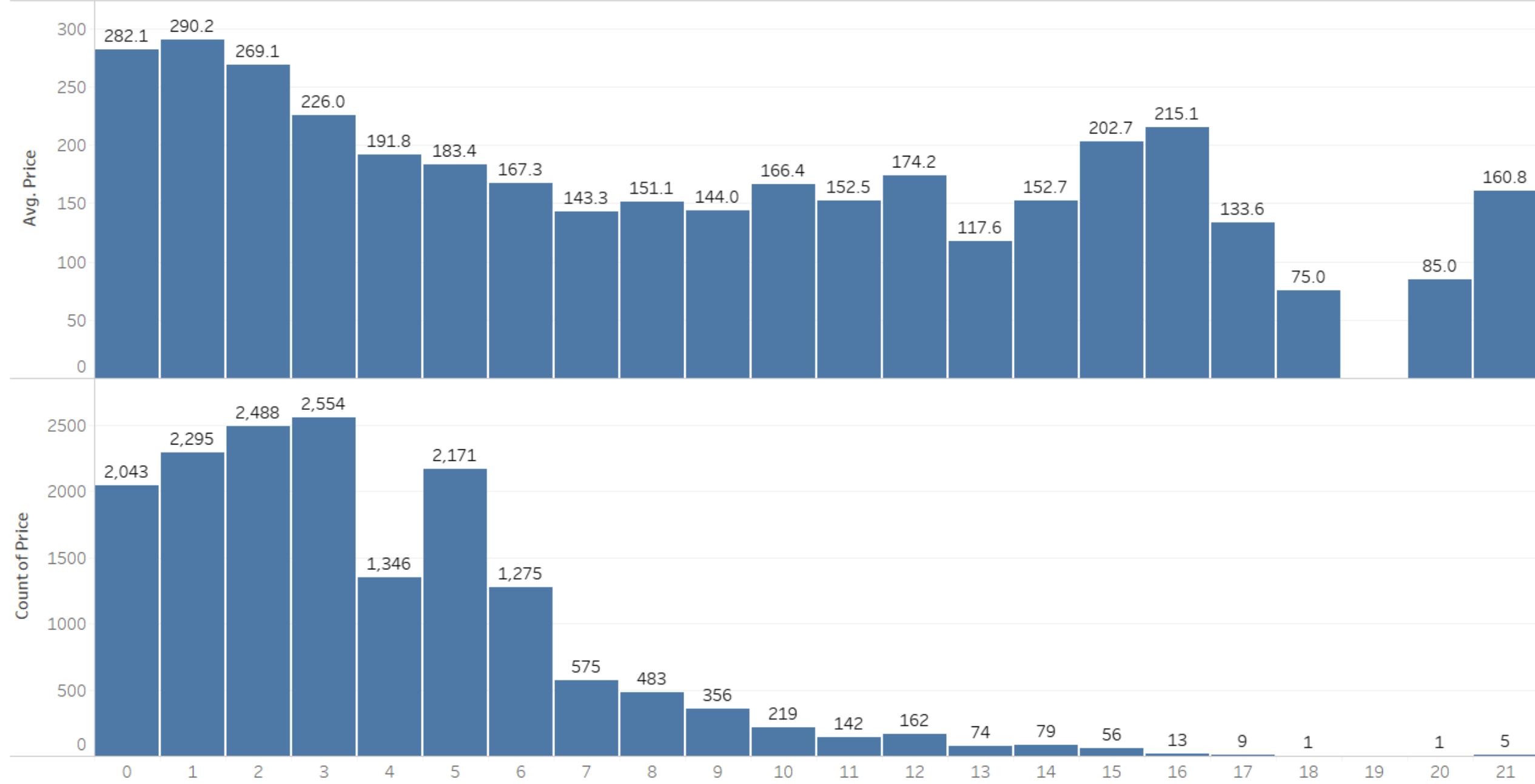
Entire House or Apartment



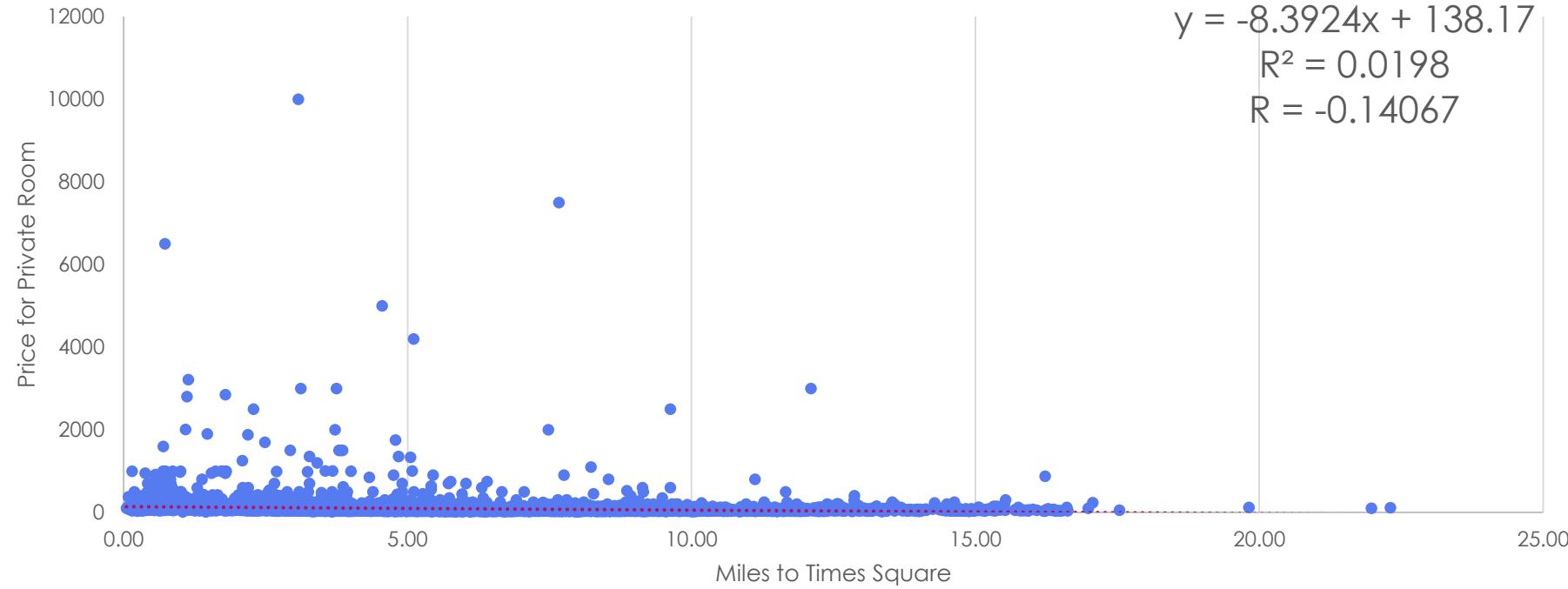
Entire House or Apt / Miles to Times Square



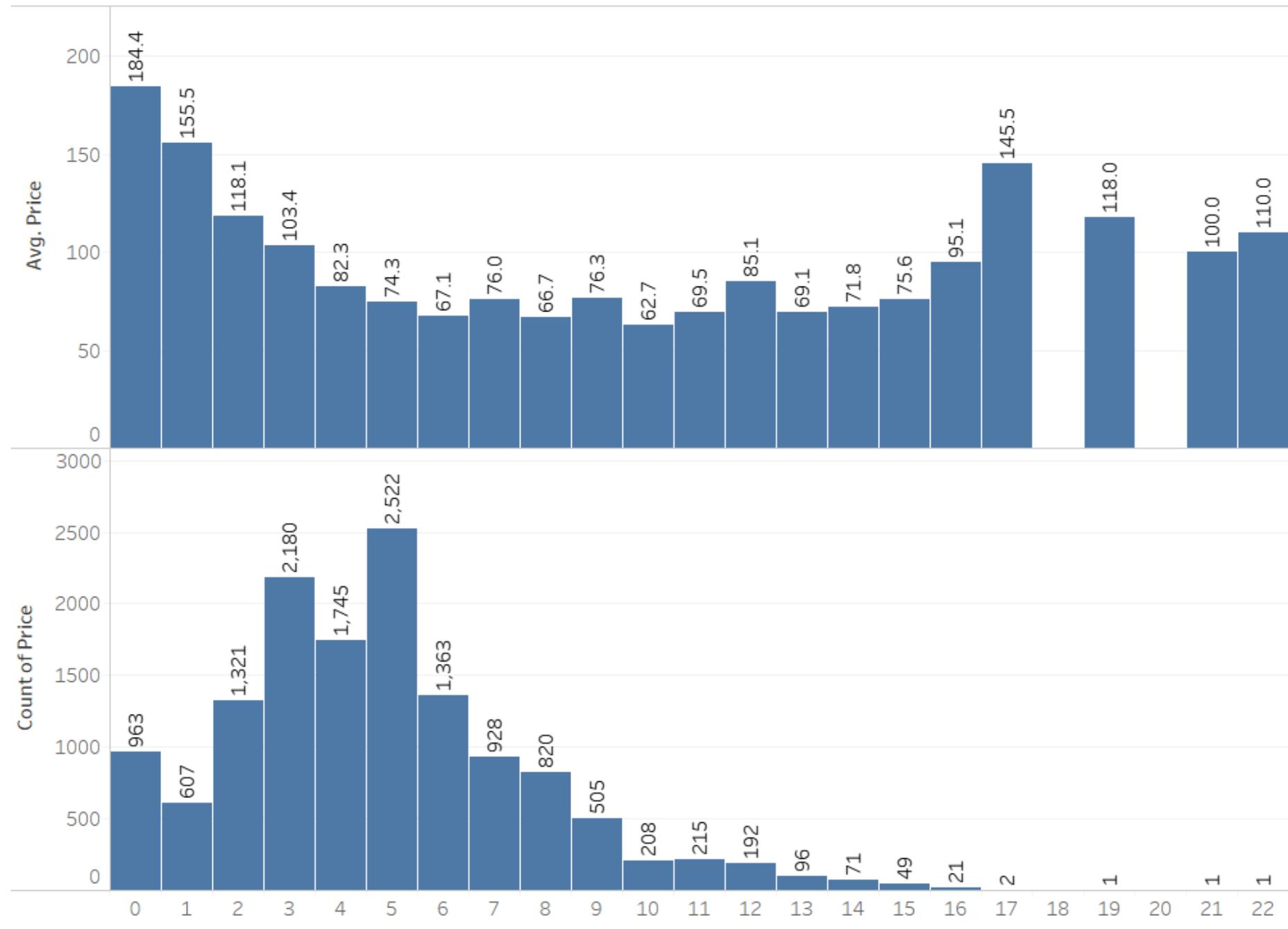
Entire House or Apt/ Miles to Times Square (with exception)



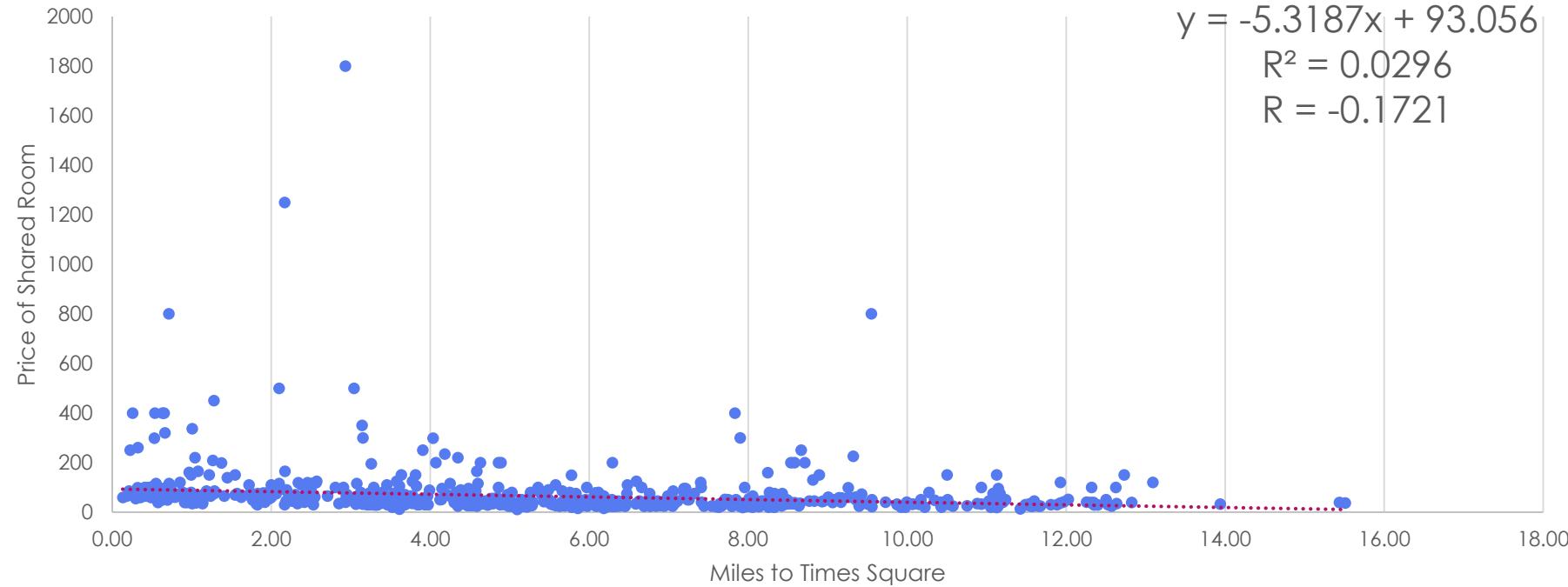
Private Rooms



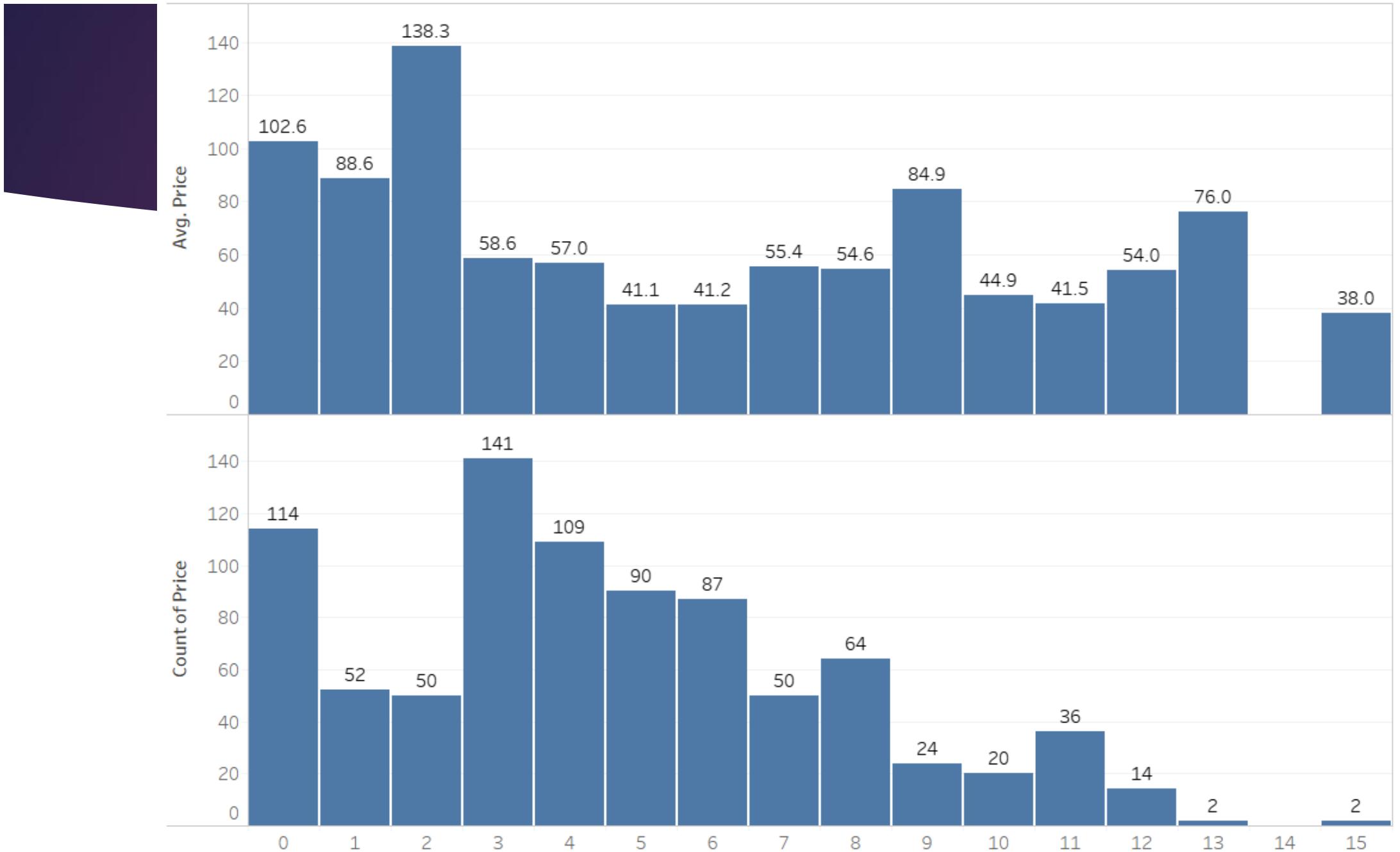
Private Room / Miles to Times Square



Shared Room



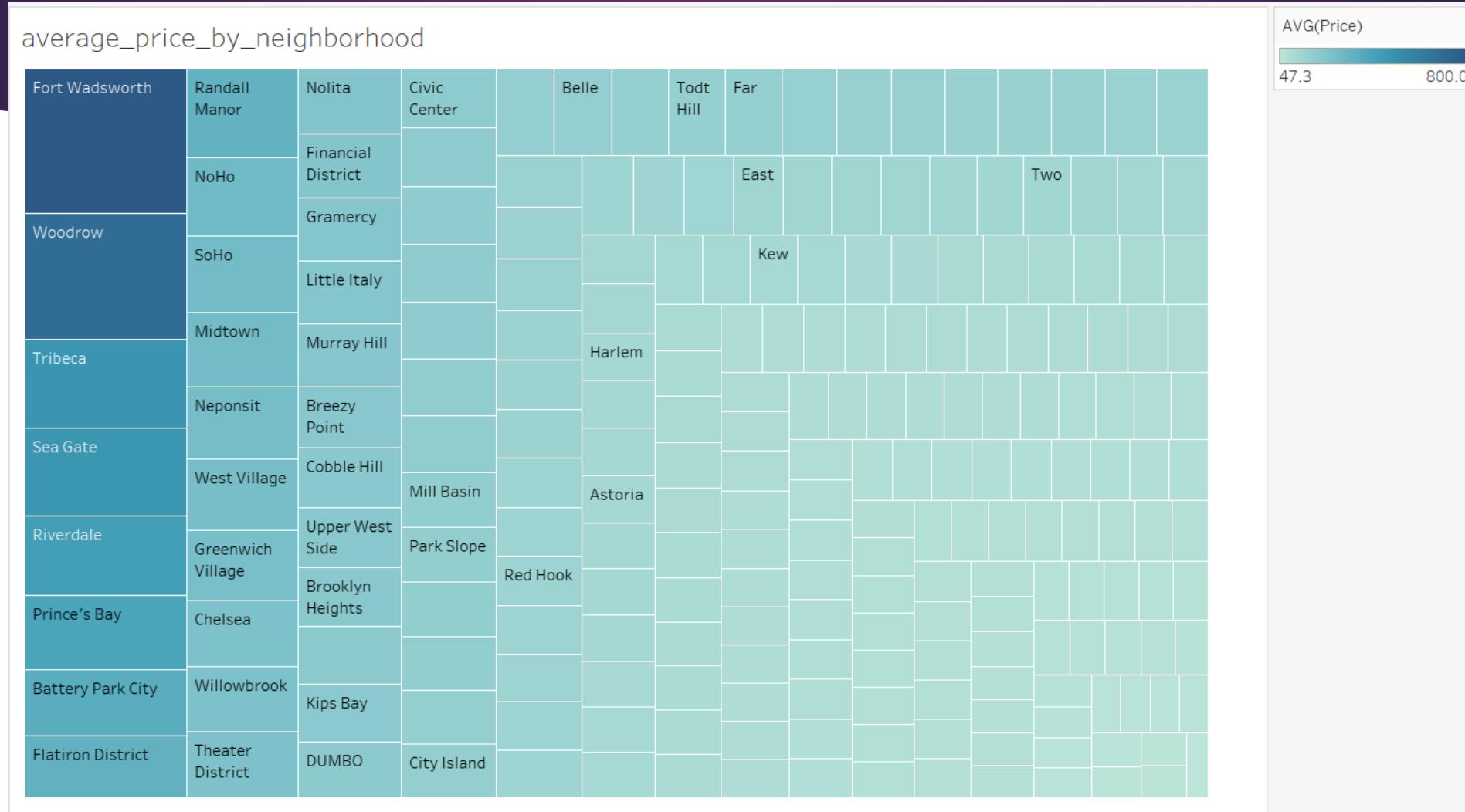
Shared Room / Miles to Times Square



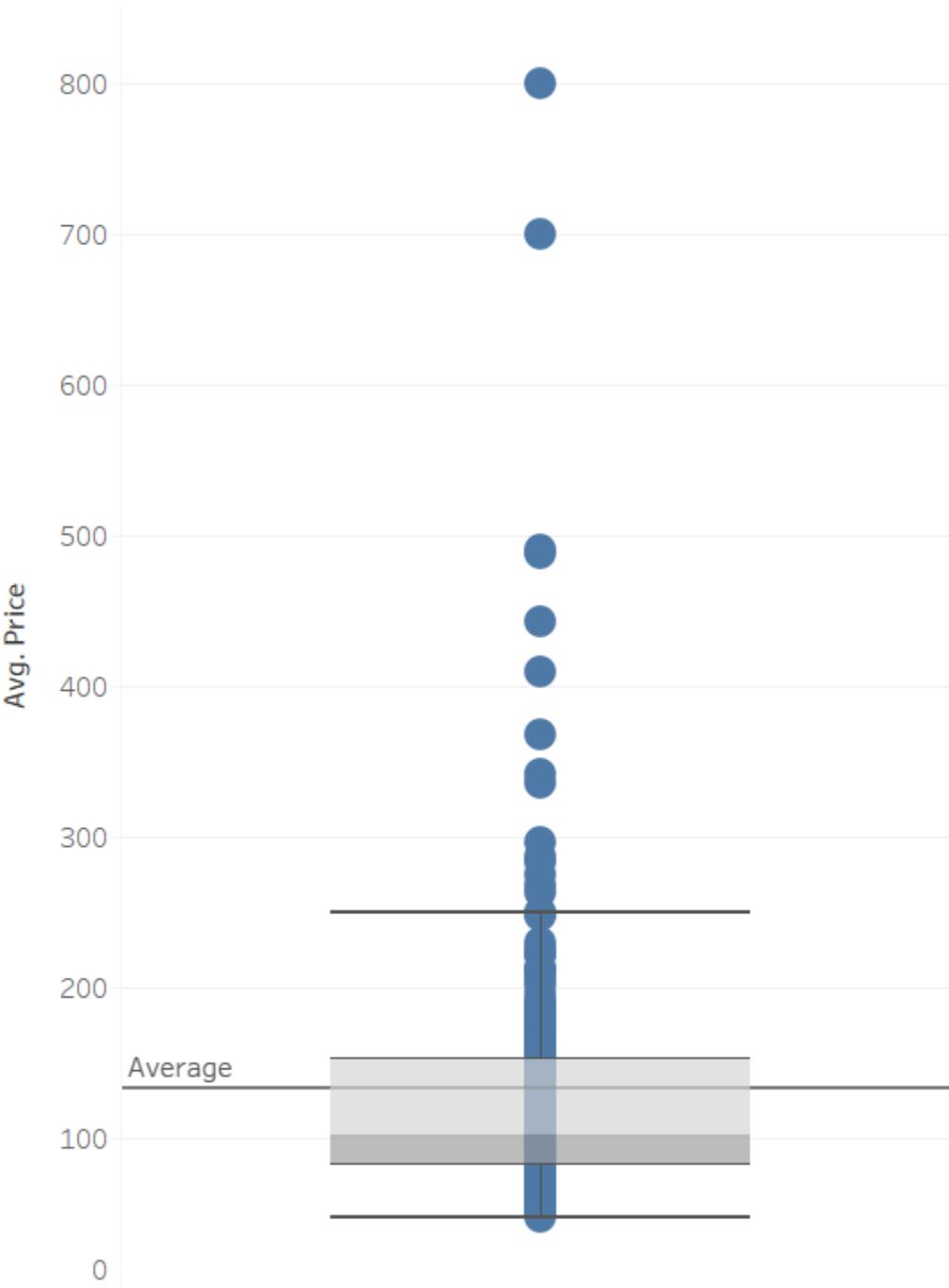
Appendix

- ▶ A bunch of maps and descriptive statistics

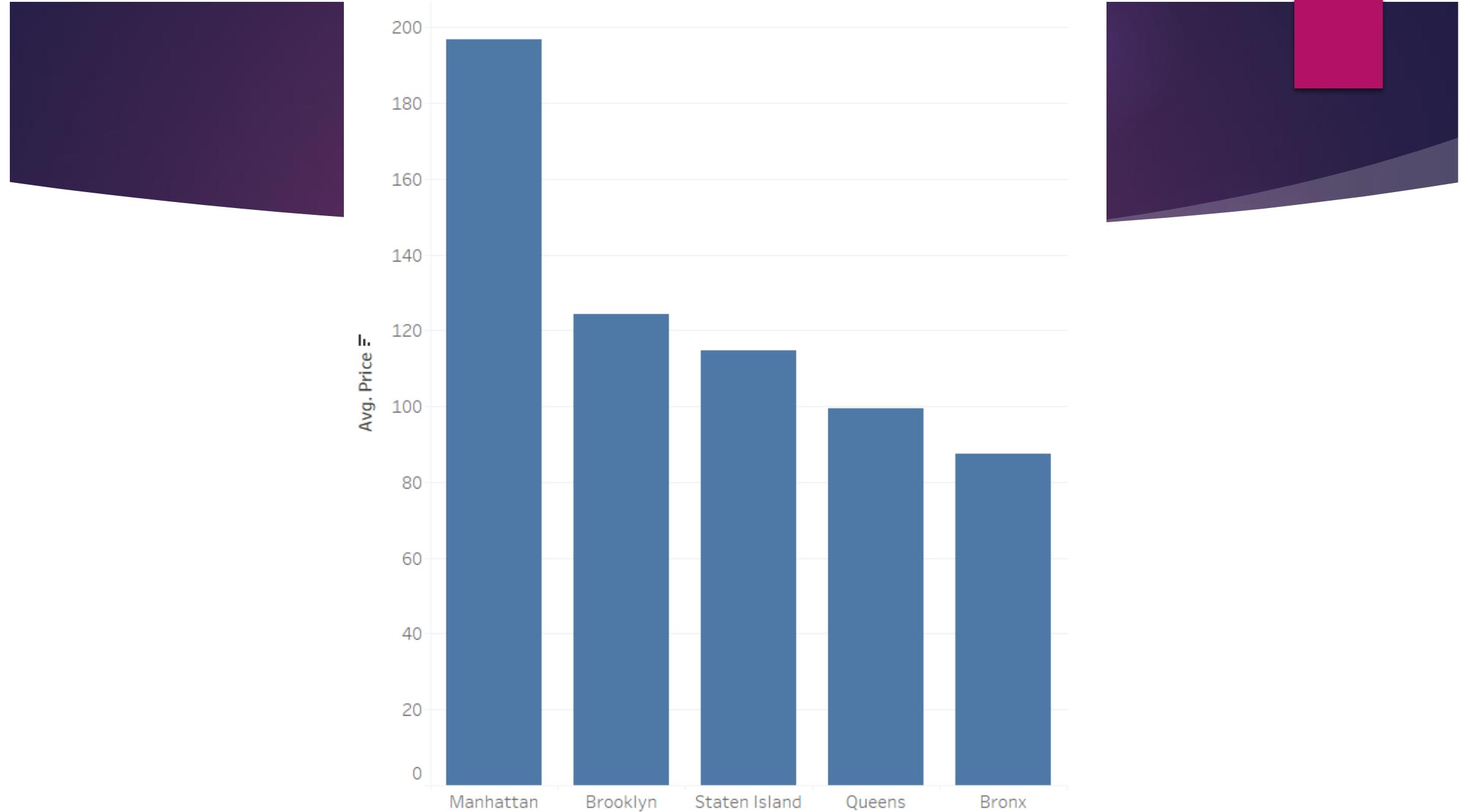
The basics



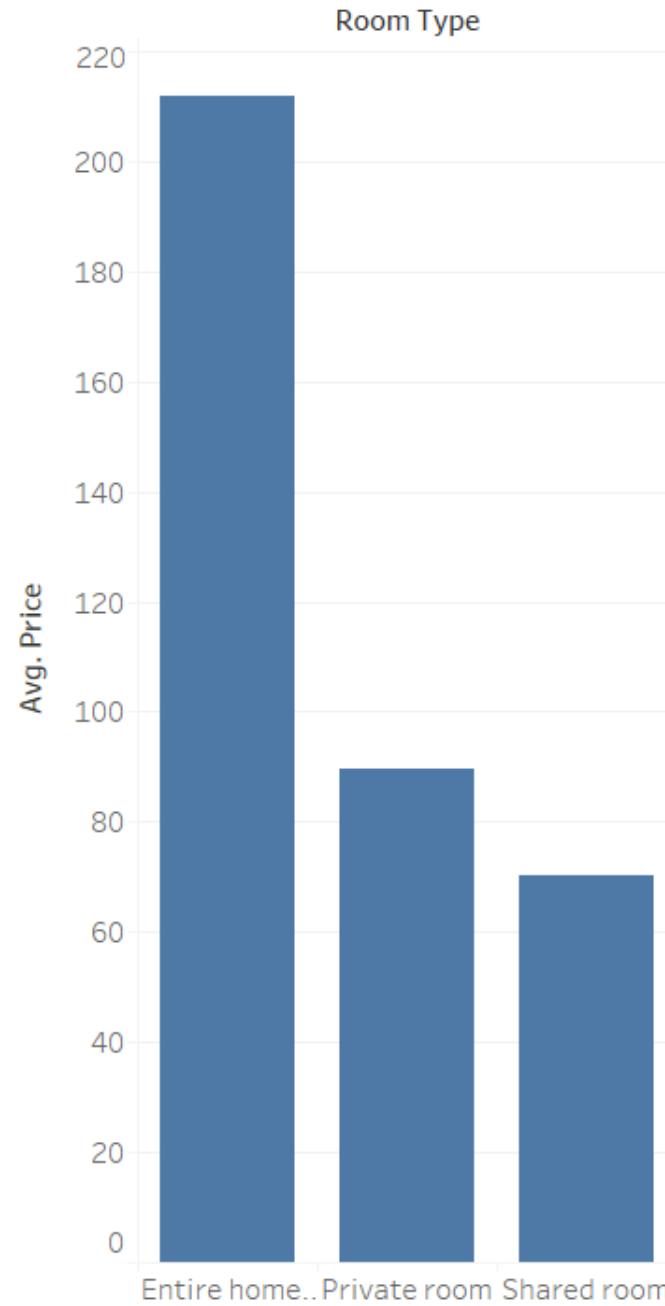
Outliers Boxplot



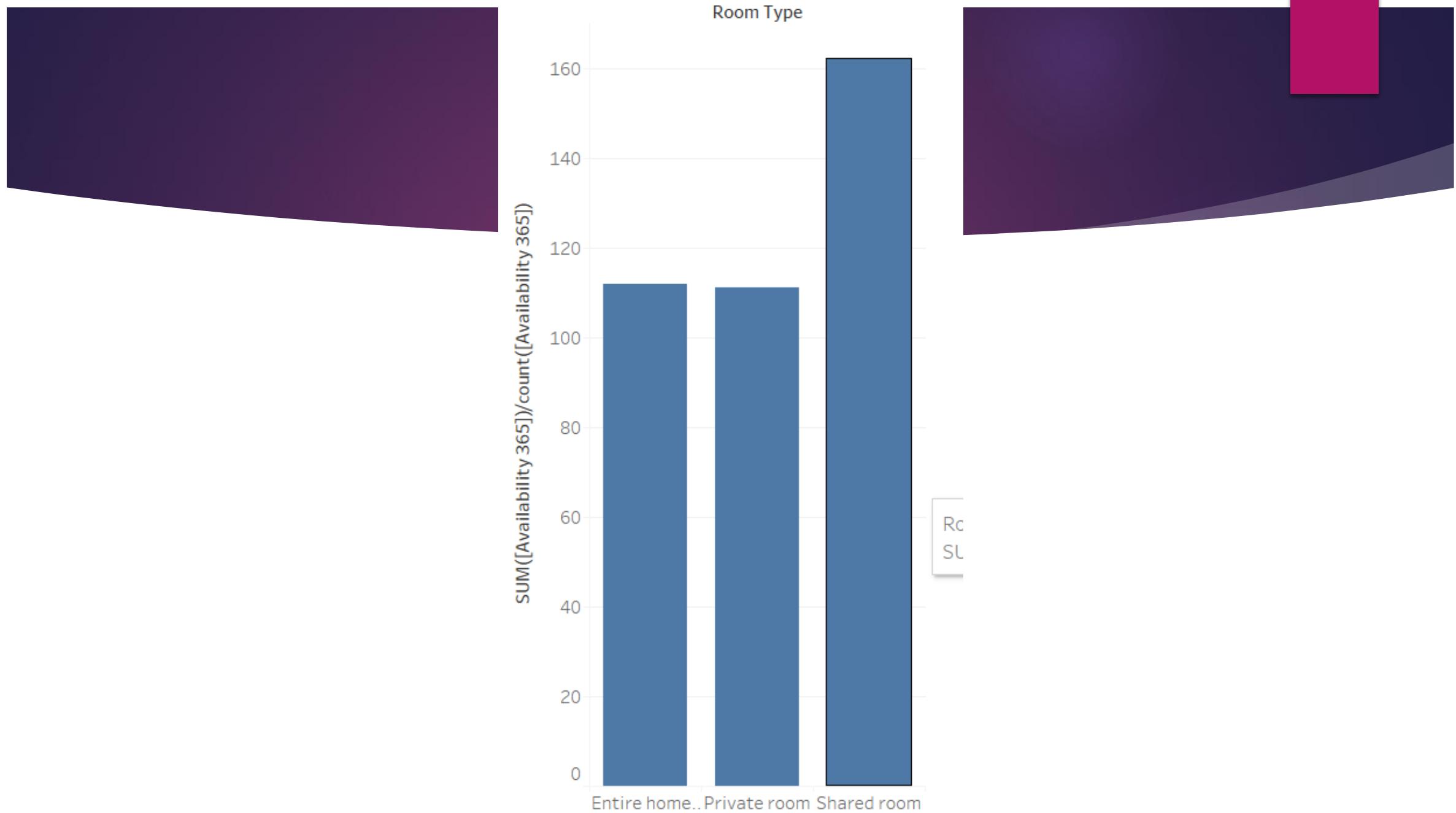
<Average Price by Borough>



AVG Price by Room Type



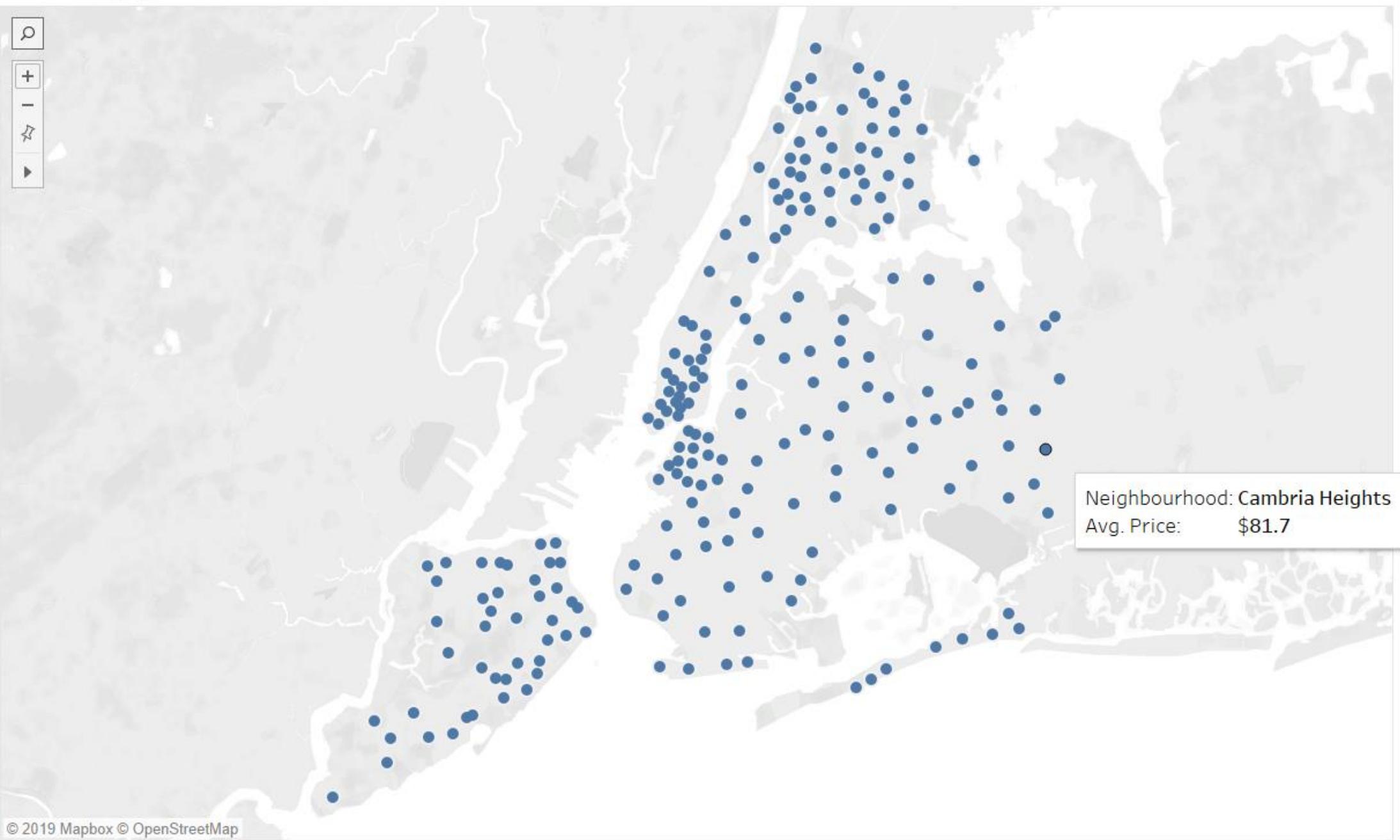
AVG Availability by Room Type





Neighbourhood Group
Bronx
Brooklyn
Manhattan
Queens
Staten Island

Sheet 6 (2)



- Neighbourhood Group
- (All)
 - Bronx
 - Brooklyn
 - Manhattan
 - Queens
 - Staten Island

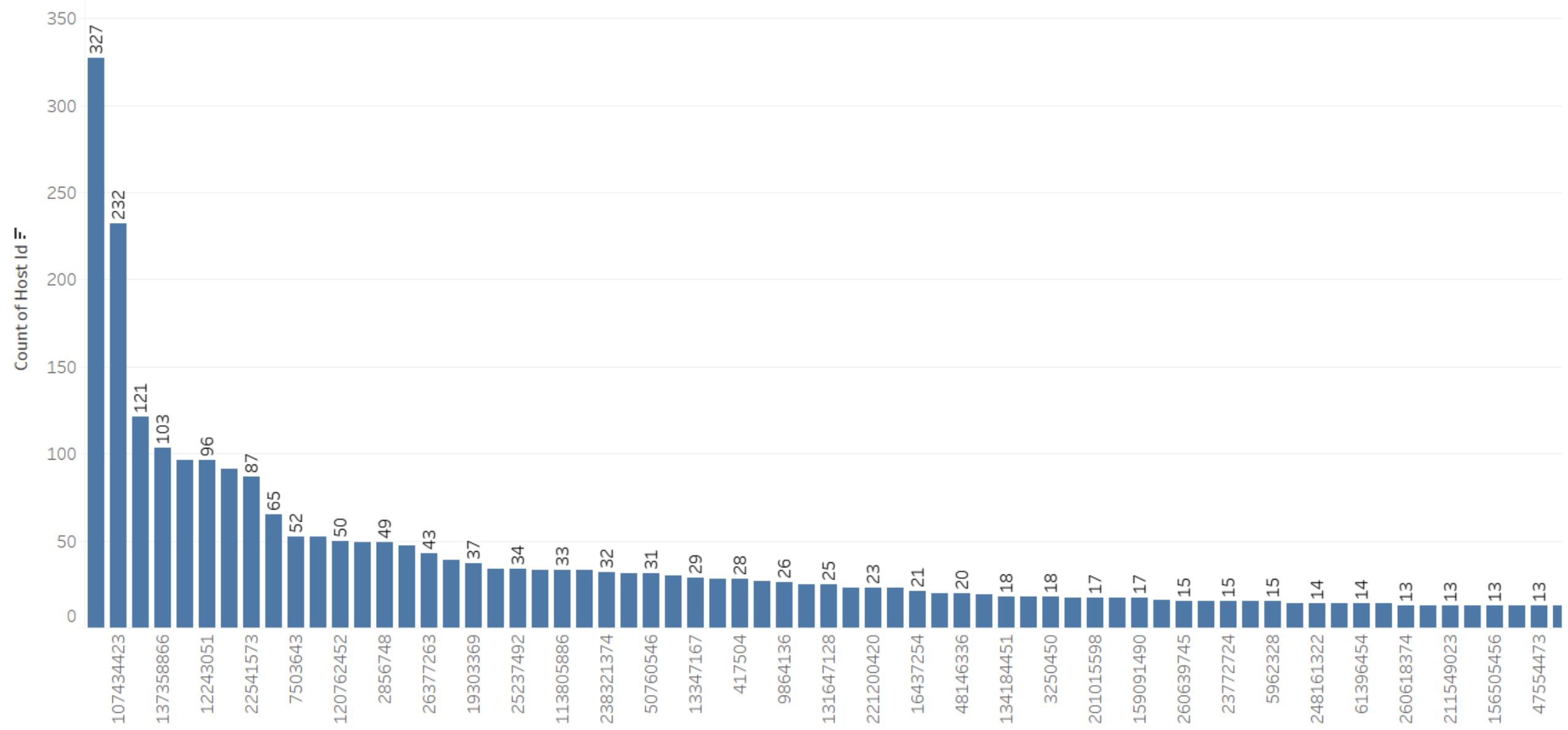
Price Slider



Number of Rentals



Host Id



```
1 SELECT COUNT (DISTINCT latitude)
2 FROM new_york WHERE host_id = 219517861;
3 |
```

Data Output Explain Messages Notifications

	count	
	bigint	
1	245	

Sonder is an AirBnB competitor startup?

- ▶ The Airbnb competitor says it tripled the number of rentable units in its online marketplace last year with more than 8,500 spaces in 20 cities around the world available today. San Francisco-based Sonder expects a \$400 million revenue run rate by the end of 2019, representing 4x growth year-over-year.

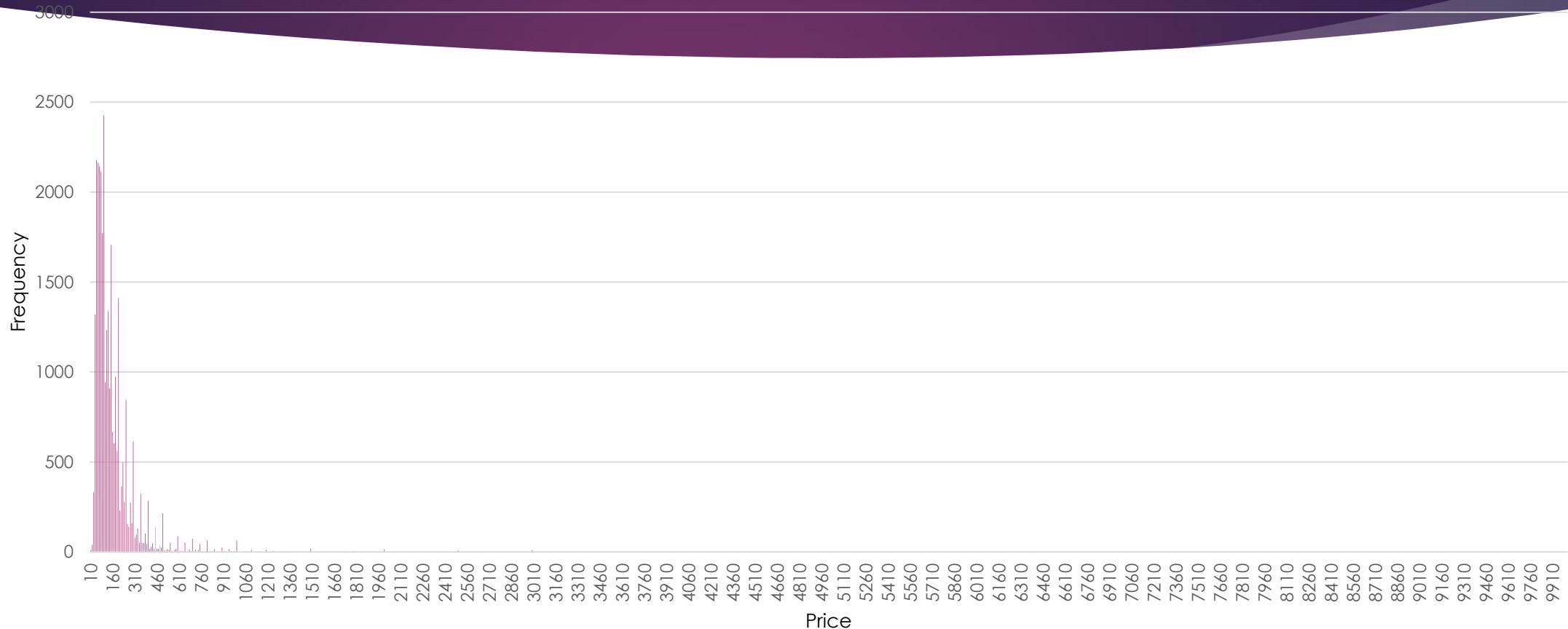
Blueground rental startup

- ▶ Blueground, the startup providing turnkey flexible rental apartments, has raised \$20 million in a round led by Athens-based VentureFriends, with participation from Endeavor Catalyst, Dubai's Jabbar Internet Group and serial entrepreneur Kevin Ryan. Ryan — who helped found MongoDB, Gilt Groupe, Zola and others — will also join Blueground's board of directors.

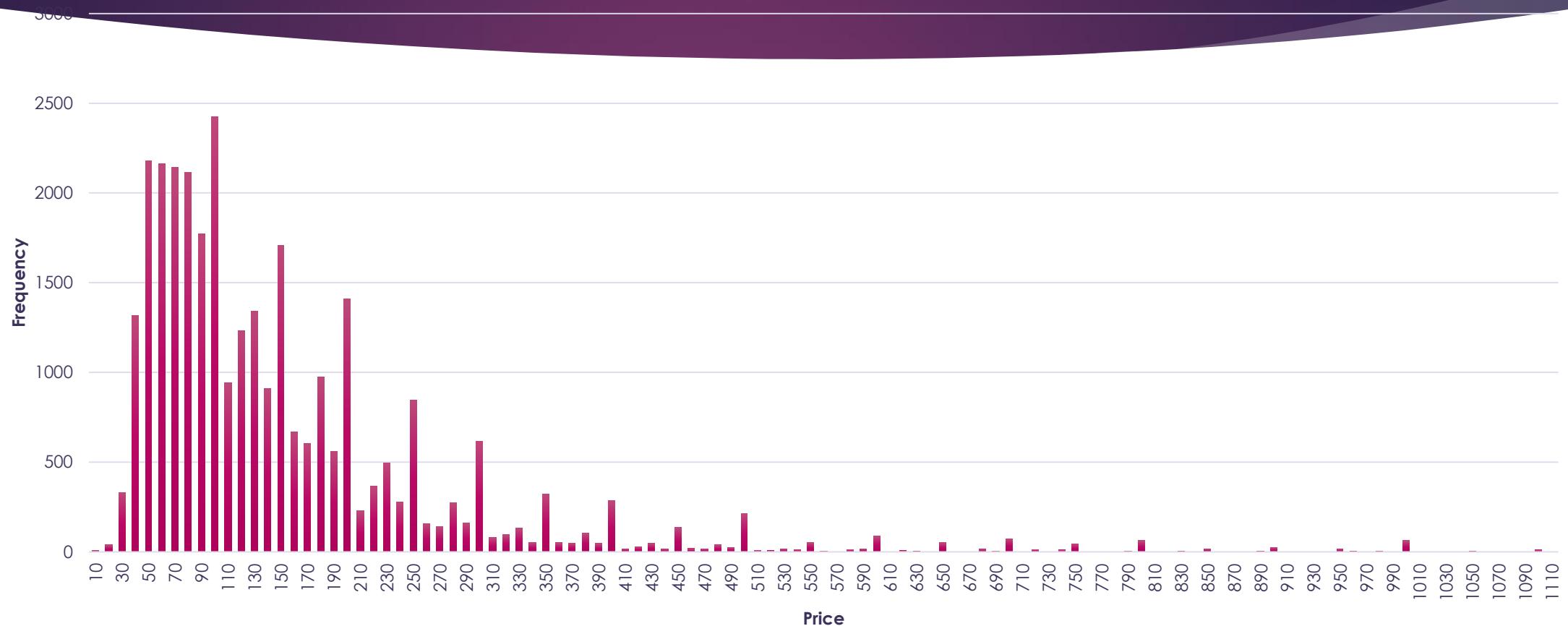
The top renters are corporations not individuals

Talk about removing the Zeroes, in
price and availability

Yeah, but is it normal

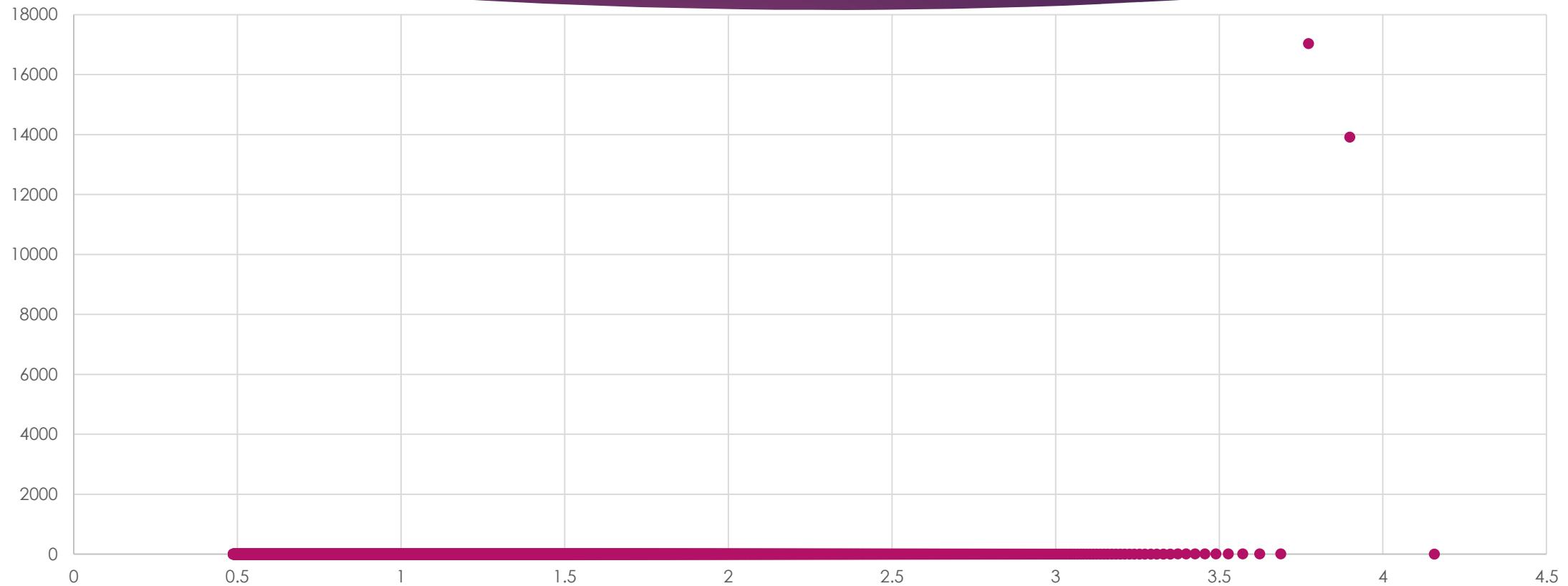


Is it though?



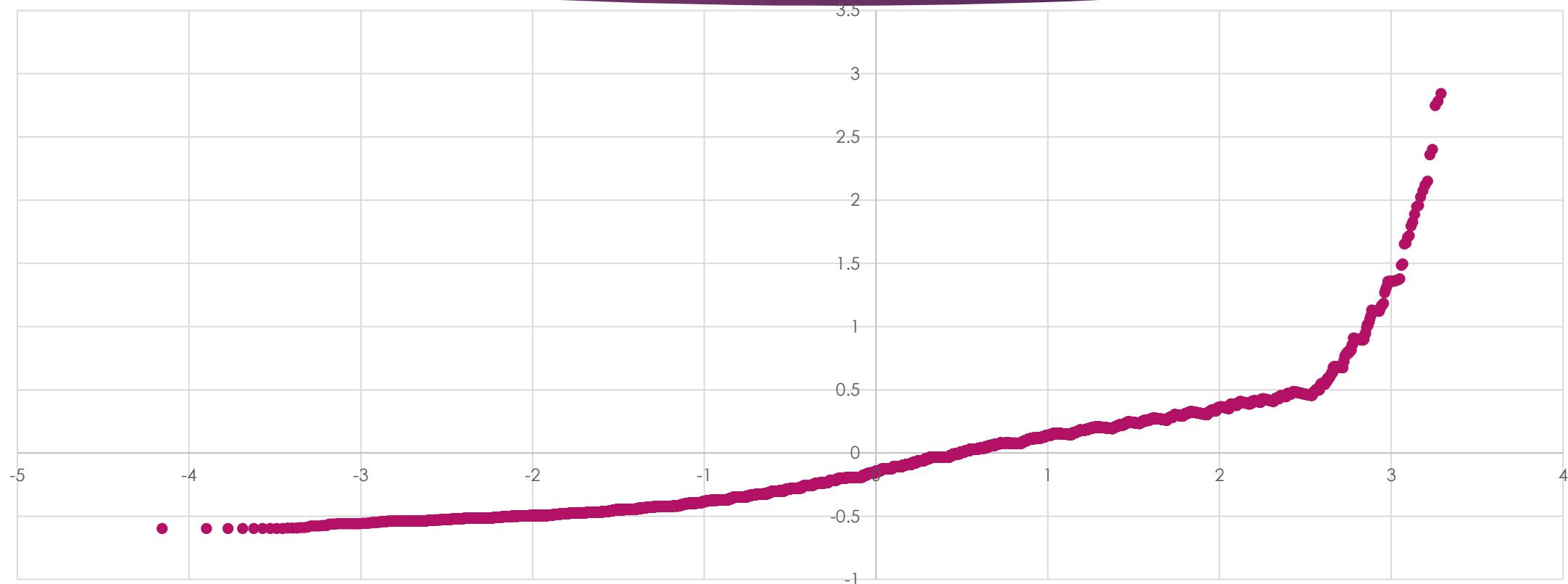
QQ Plot, 31014 Listings

Standardized X



QQ Plot, 31000 Listings

Standardized X



QQ Plot, 30850 Listings

Standardized X

$$y = 0.2331x - 0.1277$$

R² = 0.9721

