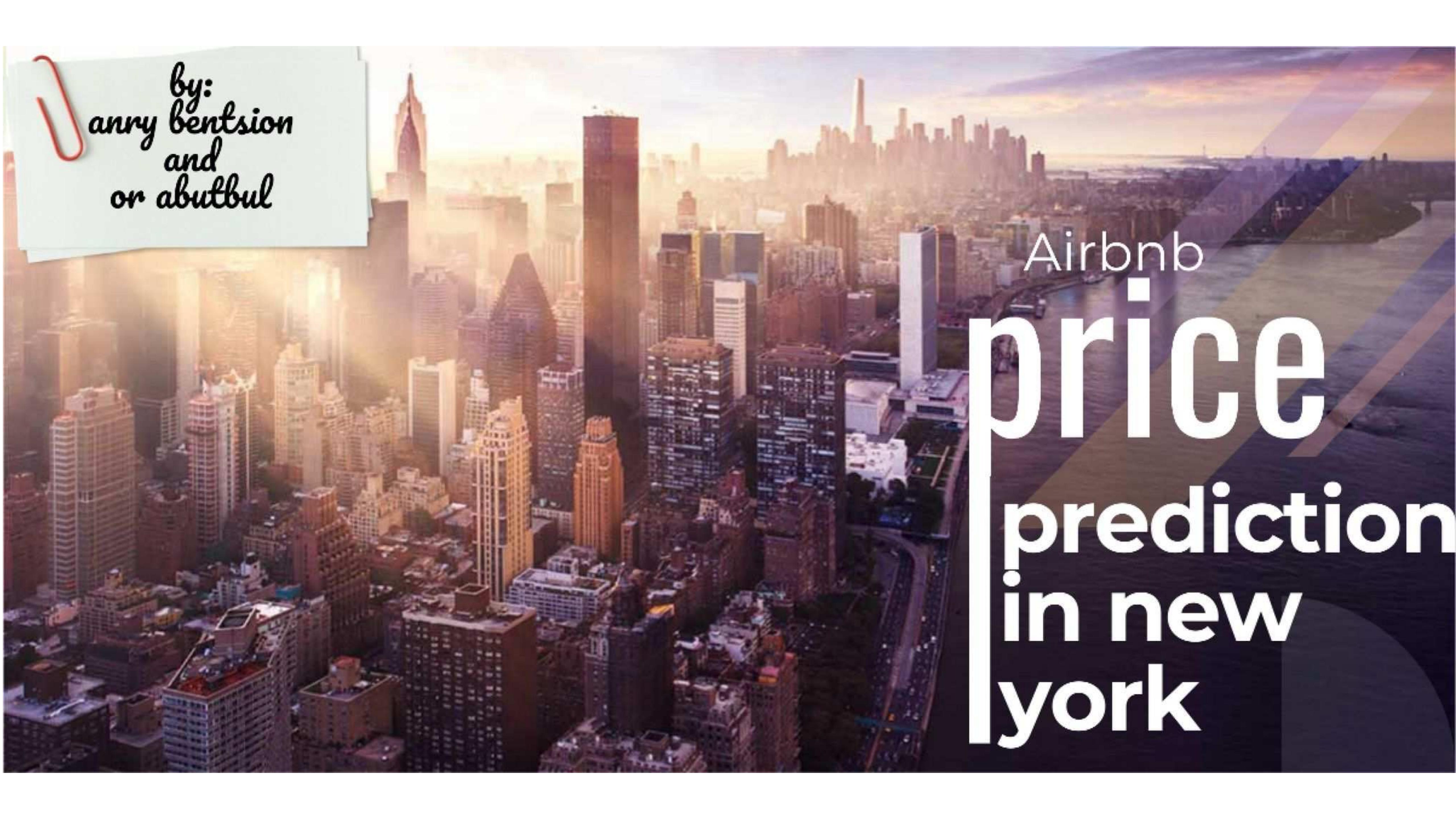




by:  
Janry Bentzion  
and  
Or Abutbul



Airbnb  
**price  
prediction  
in new  
york**

# A LITTLE BACKGROUND:

AIRBNB PROVIDES A PLATFORM FOR HOSTS TO ACCOMMODATE GUESTS WITH SHORT-TERM LODGING AND TOURISM-RELATED ACTIVITIES. GUESTS CAN SEARCH FOR LODGING USING FILTERS SUCH AS LODGING TYPE, DATES, LOCATION, AND PRICE,



GUESTS CAN ALSO SEARCH FOR SPECIFIC TYPES OF HOMES, SUCH AS BED AND BREAKFASTS, UNIQUE HOMES, AND VACATION HOMES. HOSTS PROVIDE PRICES AND OTHER DETAILS FOR THEIR RENTAL OR EVENT LISTINGS, SUCH AS THE ALLOWED NUMBER OF GUESTS, HOME TYPE, RULES, AND AMENITIES.

# RESEARCH QUESTION:

CAN WE PREDICT THE PRICE OF A RENTAL BASED ON LOCATION & TYPE ?



# PROJECT OVERVIEW:

Main steps of the project:

## Scrubbing Data

Cleaning,  
formatting,  
and filtering  
the data



## Exploring Data

Visualizing  
and  
understanding  
the data



## Modeling Data

divide the data in  
to two groups ,  
modeling different  
algorithms .



## Interpreting Data

understanding  
and  
delivering the  
results + best  
predict model



# DATA ACQUISITION :

**Unfortunately , Airbnb Isn`t allowing new user registration to their API service . Also , we found that they block many important fetures about a property in their website - like exact location (they show a radius on a map) , yearly availability and more . also , they have 30+ types of different lodging types , some only have 2-3 rentals available . therefore we found a Dataset which contains info about Airbnb rentals in the city of New York at 2019 .**

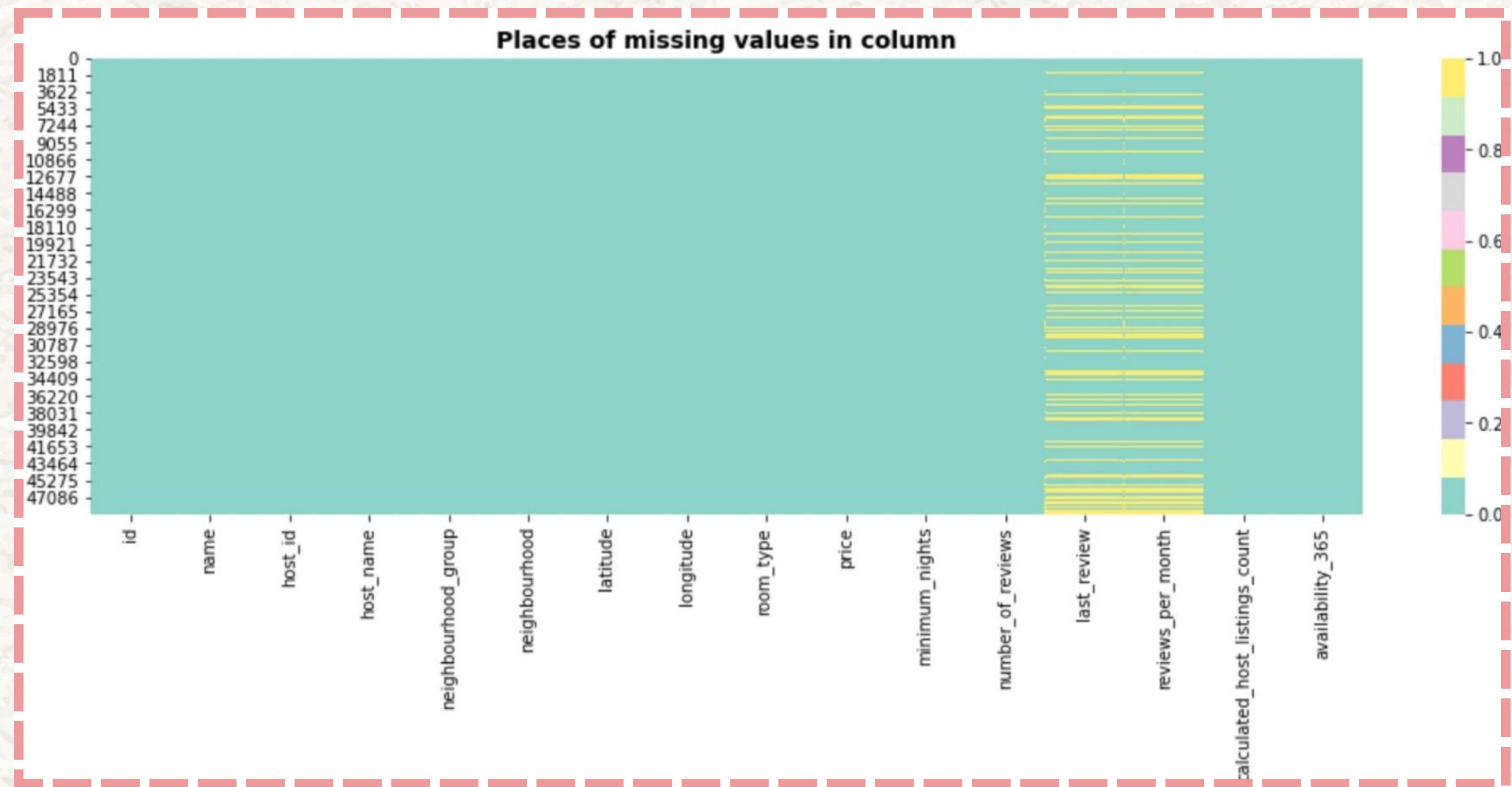
# DF. HEAD:

	<b>id</b>	<b>name</b>	<b>host_id</b>	<b>host_name</b>	<b>neighbourhood_group</b>	<b>neighbourhood</b>	<b>latitude</b>	<b>longitude</b>	<b>room_type</b>	<b>price</b>	<b>minimum_nights</b>	<b>number_of_rev</b>
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	
2	3647	THE VILLAGE OF HARLEM....NEW YORK!	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	
5	5099	Large Cozy 1 BR Apartment In Midtown East	7322	Chris	Manhattan	Murray Hill	40.74767	-73.97500	Entire home/apt	200	3	
6	5121	BlissArtsSpace!	7356	Garon	Brooklyn	Bedford-Stuyvesant	40.68688	-73.95596	Private room	60	45	
7	5178	Large Furnished Room Near B'way	8967	Shunichi	Manhattan	Hell's Kitchen	40.76489	-73.98493	Private room	79	2	

**Num of features: 16**  
**Num of entries: 48895**

# DATA SCRUBBING :

We started with dealing with missing values , some important to our analysis and some dont . we dropped the columns that are insignificant to our research , dropped duplicates and filling NaN values .

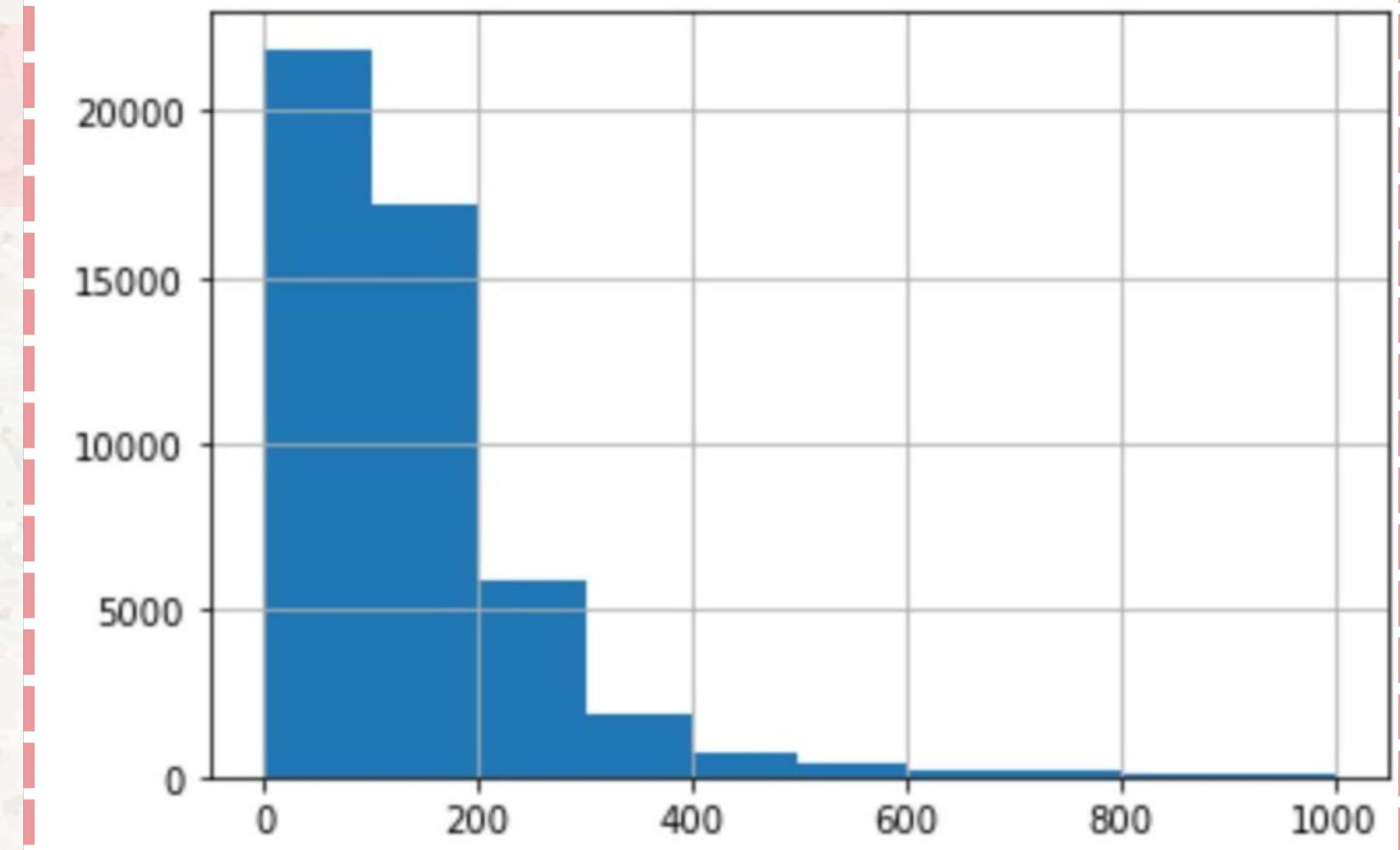


# HANDLING OUTLIERS :

Price varies between 0 to 10,000 , average price is 152 .  
we used a histogram to see how the prices scattering .  
therefore , we decide to drop all entries that cost more than 250\$ for a better modeling ability .

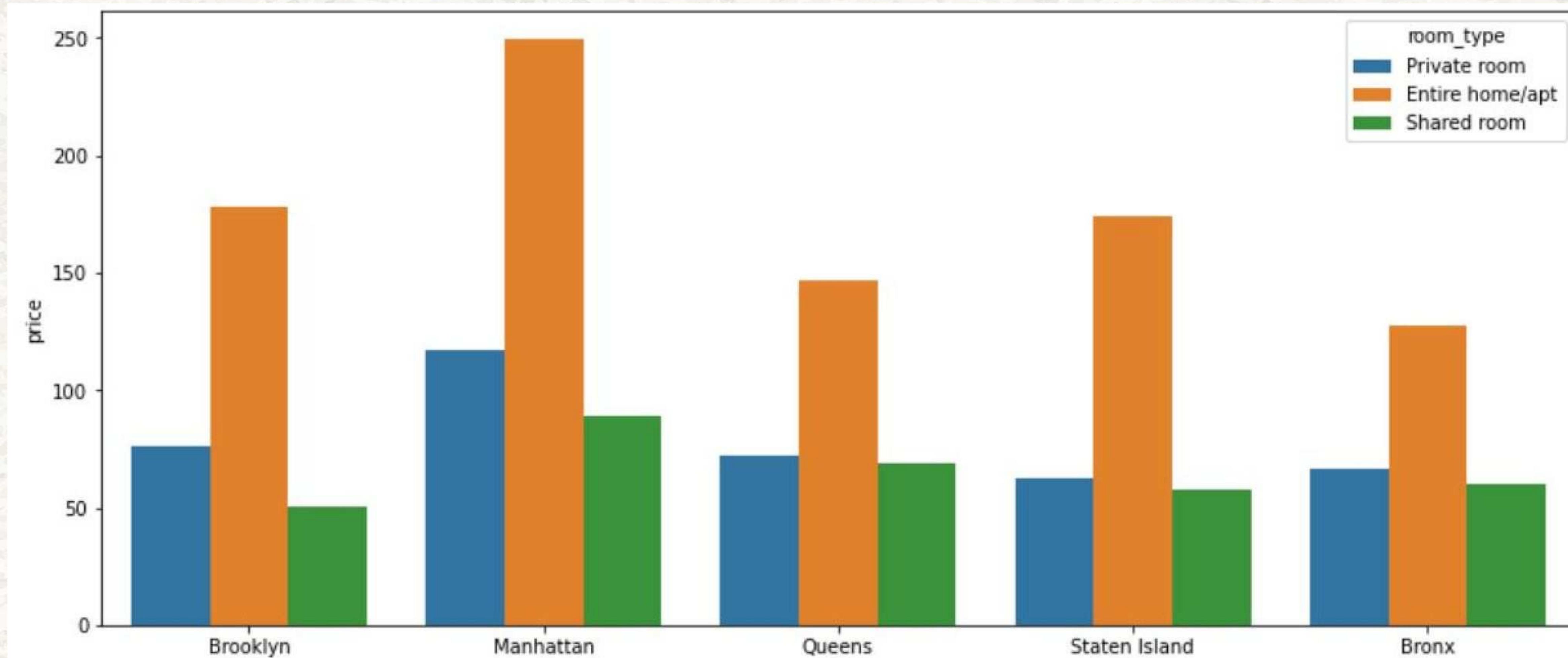
```
df['price'].describe()
```

count	48895.000000
mean	152.720687
std	240.154170
min	0.000000
25%	69.000000
50%	106.000000
75%	175.000000
max	10000.000000



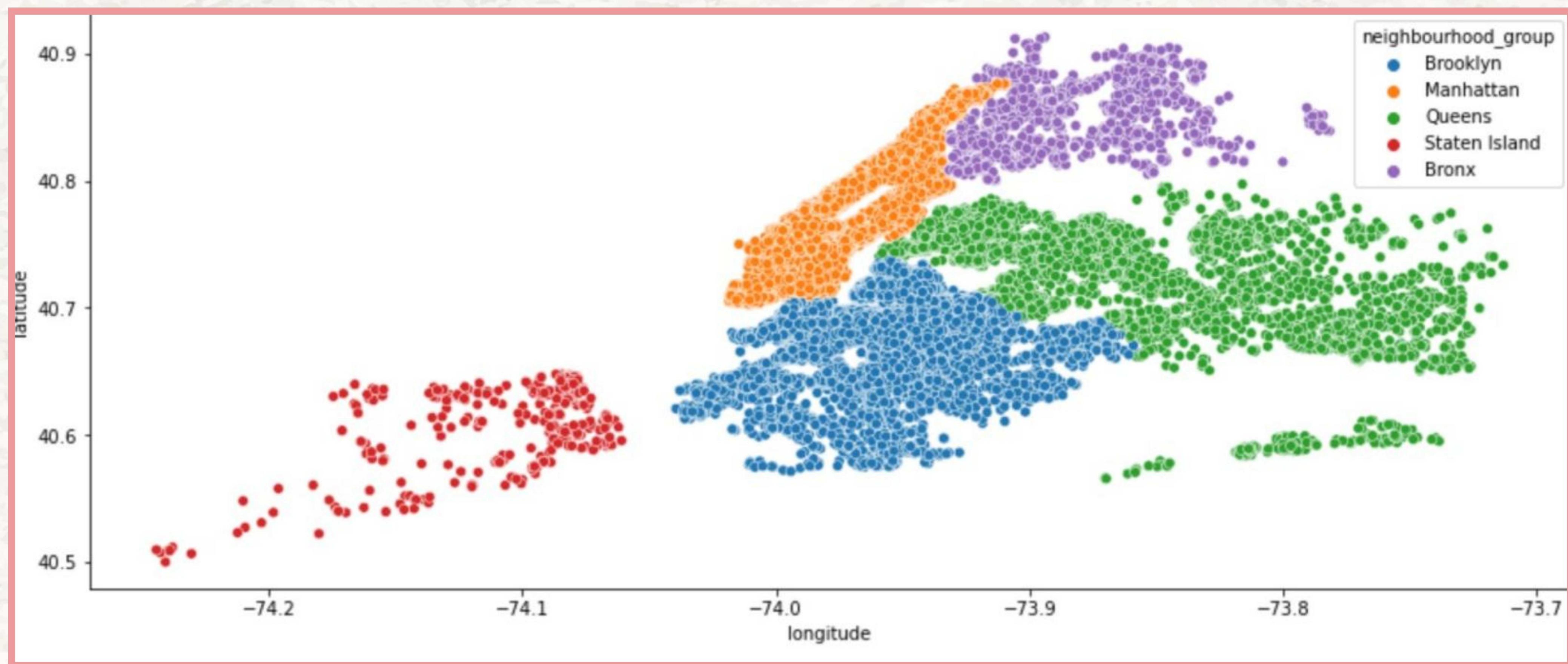
# EXPLORING & VISUALIZING THE DATA:

**we can see that Manhattan is the most expensive neighbourhood.  
Bronx is the cheapest , but if your willing to get a Shared room you should look  
for one in Brooklyn .**



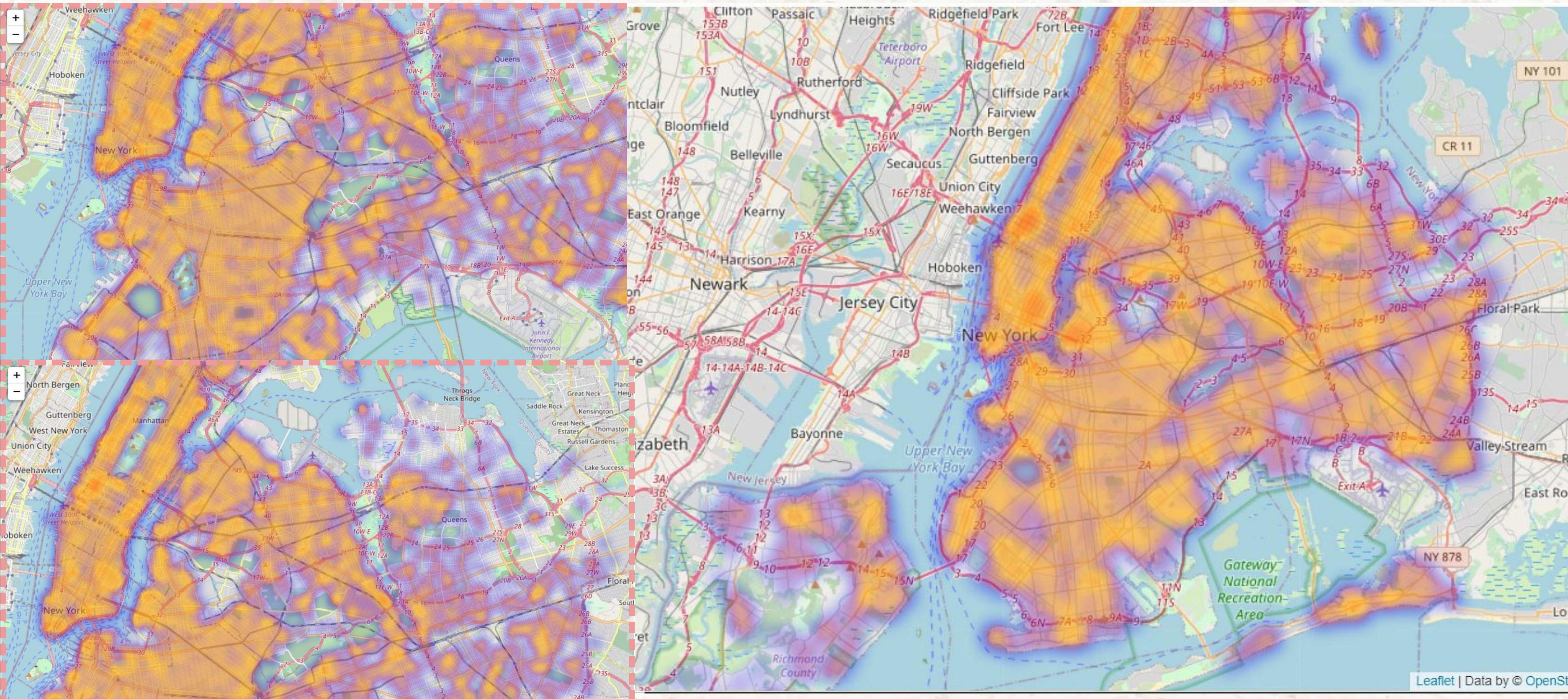
# EXPLORING & VISUALIZING:

Visualizing locations using  
scatter plot with the properties  
coordinates :



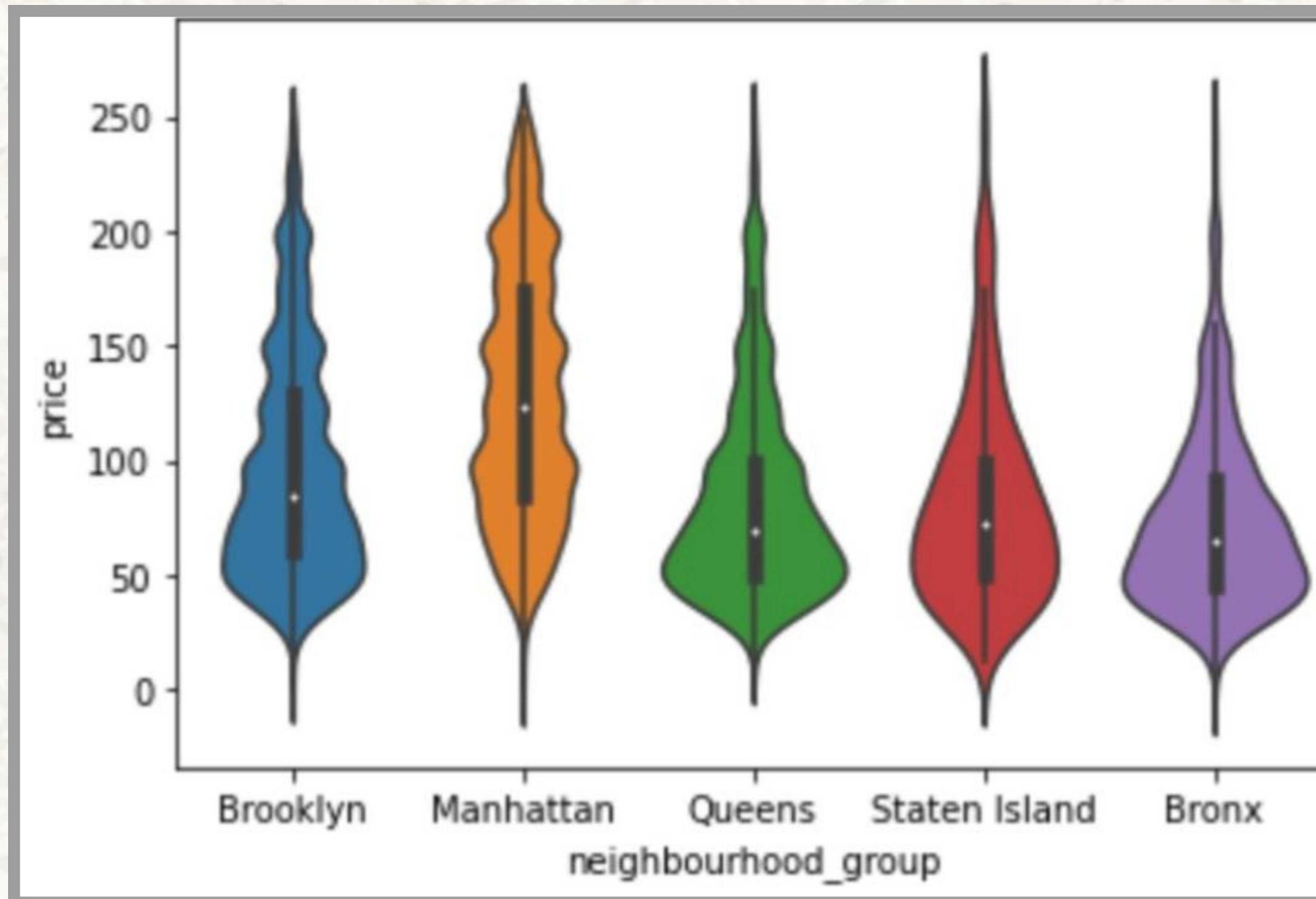
# EXPLORING & VISUALIZING:

Distribution of Airbnb rentals in a Heat Map , in regard to every rentals price using Folium :

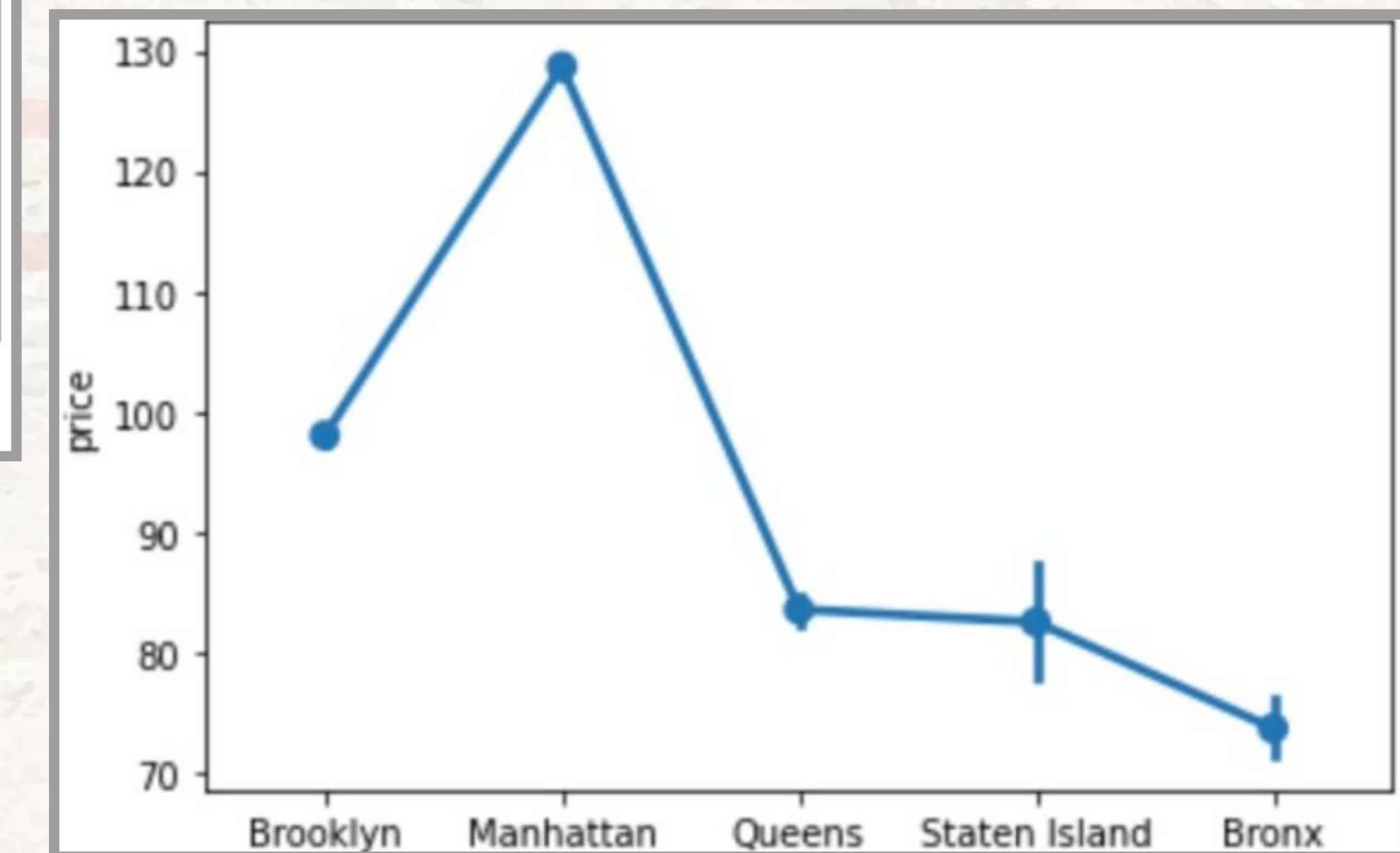


# EXPLORING & VISUALIZING:

Distribution of Airbnb rentals in a violinplot :



Average price for each neighbourhood using pointplot :



# EXPLORING & VISUALIZING:

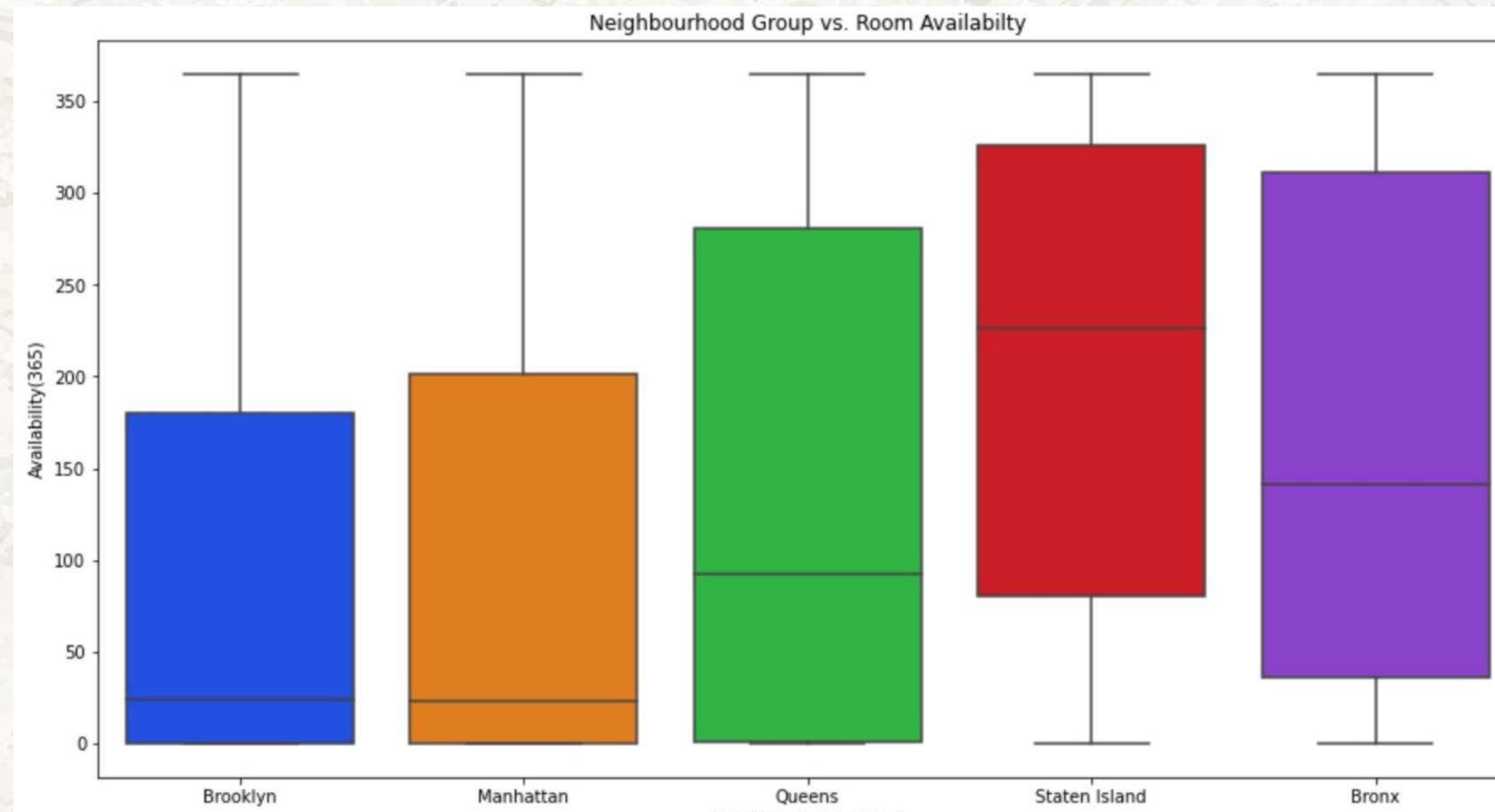
Check for correlation between different features :



Our features are not coorelated.

# EXPLORING & VISUALIZING:

lets check the availability in different neighbourhoods:



**we can clearly see that Brooklyn is the most booked neighbourhood , slightly ahead of Manhattan - which explains why they are the most expensive choice in NYC.**

# MODELING DATA :

## Supervised Learning :

First , we labeled the important features to our model . than , we used StandardScaler to scale our data . and than we applied Linear Regression learning from SKLearn Library .

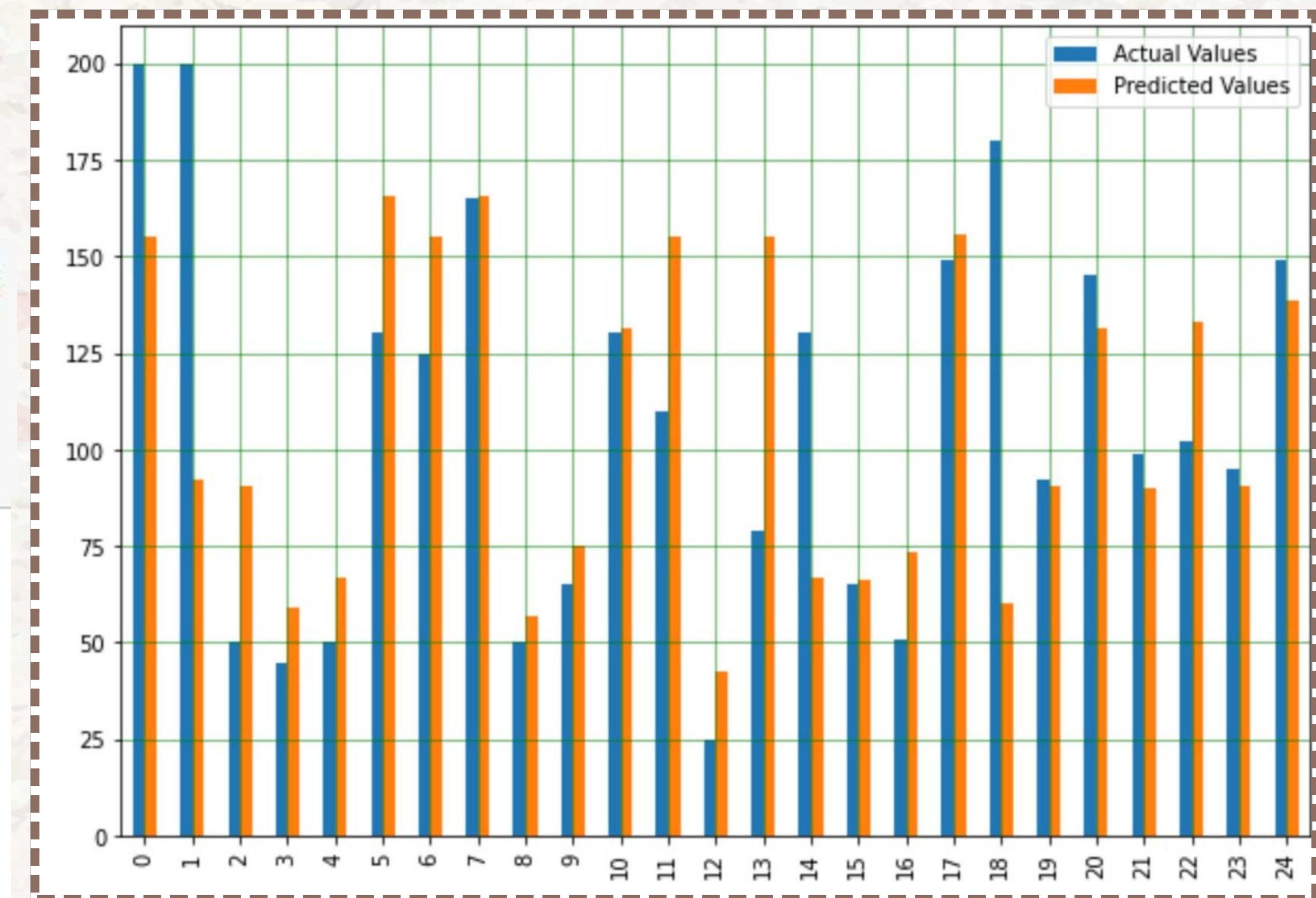
*#Prepare a Linear Regression Model*

```
reg = LinearRegression()  
reg.fit(X_train,y_train)  
y_pred = reg.predict(X_test)
```

R2 score: 0.4967021412940398

RMSE: 38.18895363010685

	Actual Values	Predicted Values
0	200	155.234454
1	200	92.199635
2	50	90.576679
3	45	59.349313
4	50	66.773859



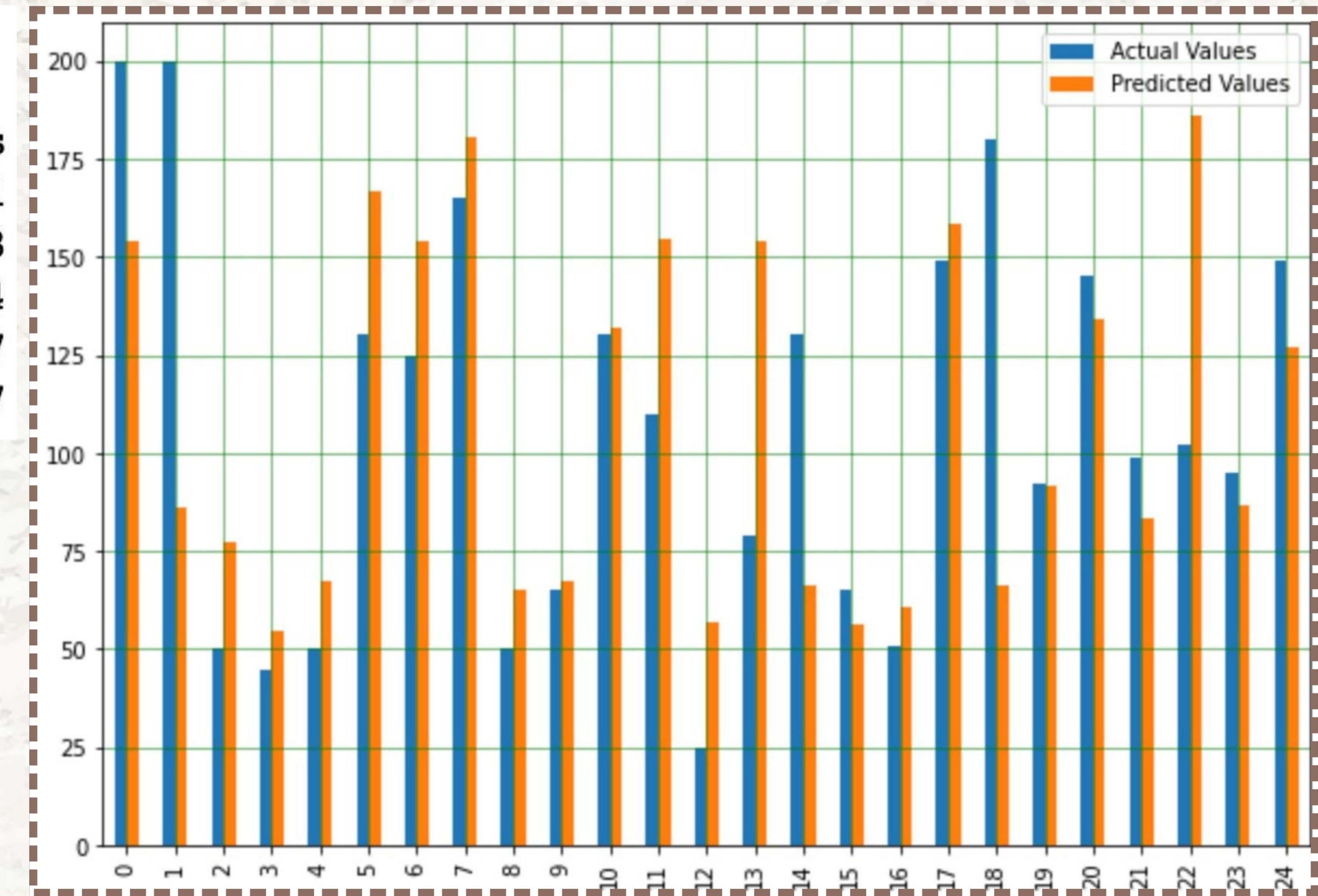
# MODELING DATA :

we also tried Random forest regression learning , it did improve the R2 score but only in 0.002 .

R2 score: 0.49825968657829733

RMSE: 38.12981656681449

	Actual Values	Predicted Values
0	200	154.023021
1	200	86.365013
2	50	77.456994
3	45	54.603957
4	50	67.167067



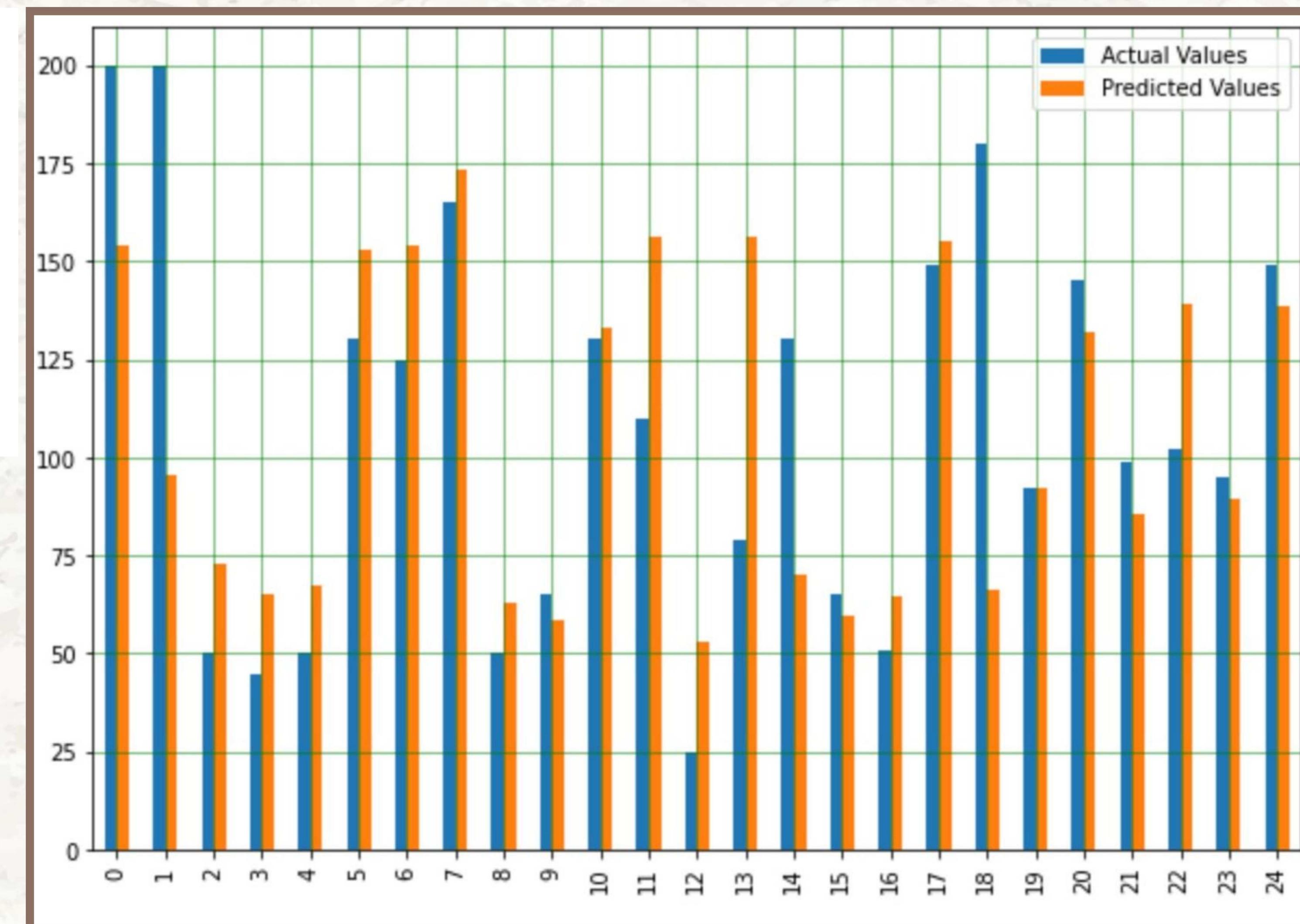
# MODELING DATA :

Finally , we changed the test size to 15% and tried the Gradient Boosting Regressor .  
that gave us the best model yet .

R2 score: 0.5236796285915105

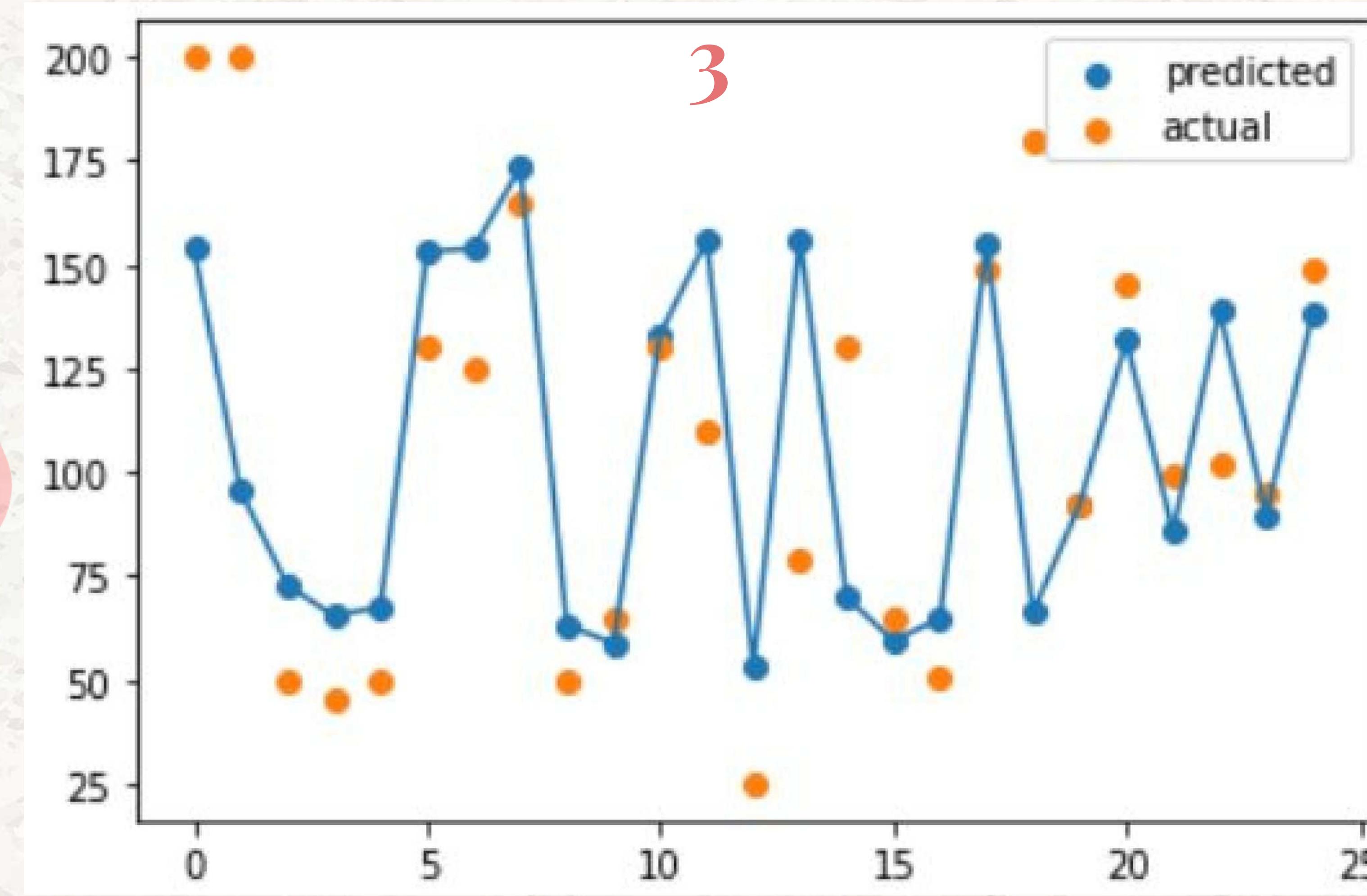
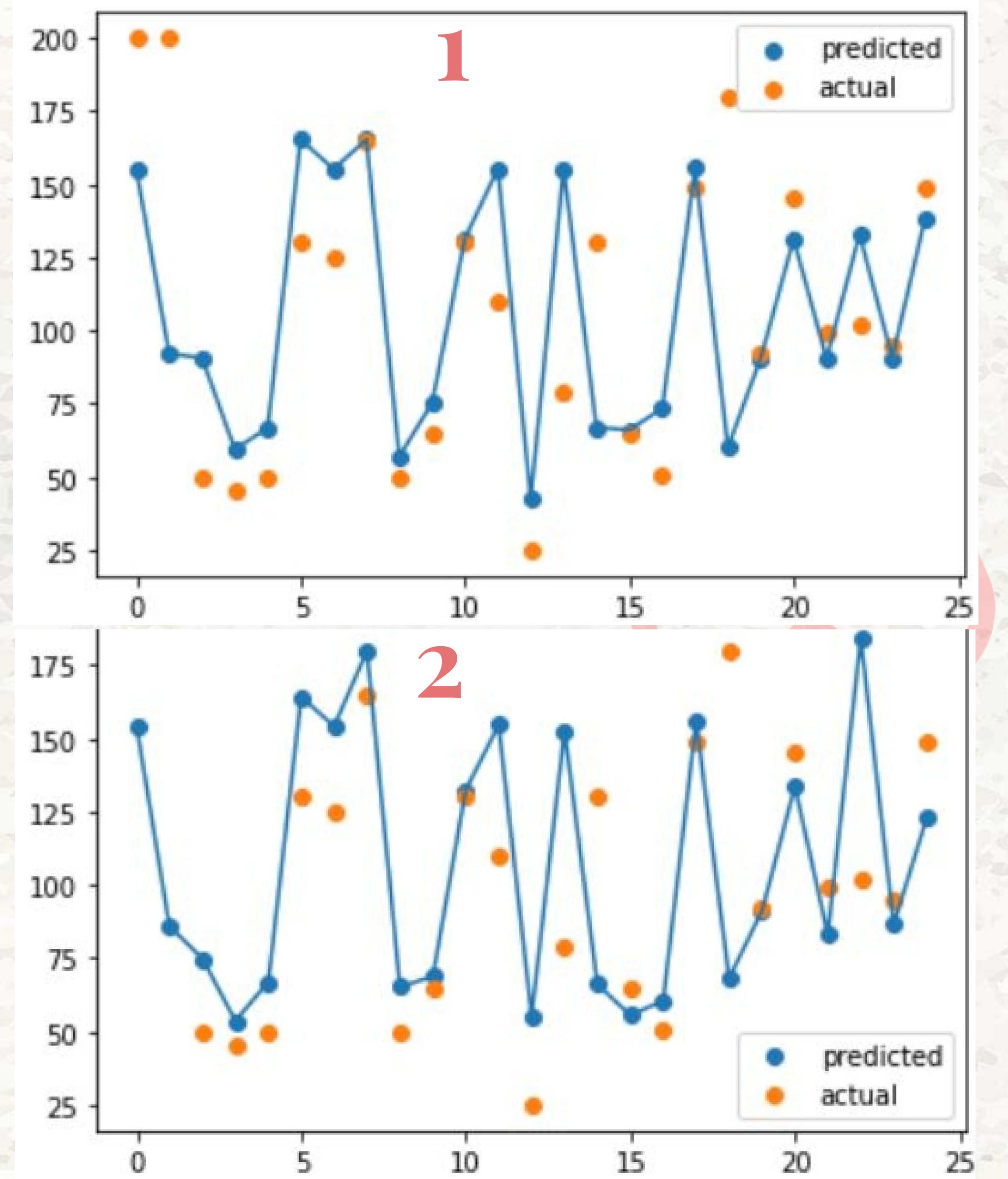
RMSE: 37.15136675572461

	Actual Values	Predicted Values
0	200	153.857779
1	200	95.504342
2	50	72.846507
3	45	65.368276
4	50	67.512499



# MODELING DATA :

Another point of view :



2

# CONCLUSION :

IN RETURN TO OUR RESEARCH QUESTION - WE CAN PREDICT THE PRICE OF A RENTAL BASED ON THE LOCATION AND THE TYPE .

WE ALSO SAW THE FOLLOWING :



TIMES CROSSING A STREET TO A DIFFERENT NEIGHBOURHOOD CAN LAND YOU A PRIVATE ROOM IN A PRICE OF A SHARED ROOM ACROSS THE STREET.

\*BROOKLYN GIVES YOU THE MOST BANG FOR YOUR BUCK IN NYC IN TERMS OF PRICE AND LOCATION , HOWEVER ITS HARDER TO FIND AN AVAILABLE PLACE THERE .

\*IF YOUR WILLING TO GO A BIT FURTHER OUT , YOU SHOULD SEARCH IN BRONX AND SAVE SOME CASH.