

Airbnb in New York City

Public Data from 2008 to 2015
Data Visualization and Tableau Dashboards

Vivien Ho



About Airbnb

Airbnb is an online company that provides a platform for finding alternative places for short-term homestays.

The company has listings all over the world where participating hosts rent out their homes and/or rooms on a wide range of pricing. Visitors can also provide a review and score of their experience at the end of their stay.



Overview

Goals:

- Understand Airbnb's New York City (NYC)'s short-term home market
- Perform exploratory data analysis to understand the relationship and trends of Airbnb's short-term homestay market in NYC.
- Create visualizations and dashboards to illustrate findings and to answer questions of concern.

Investigate:

- What is the median price for each room type in a particular neighbourhood?
- What are the room types in a particular zipcode and their respective prices and rating score?
- Which zipcodes have the highest ratings (with over 100 number of reviews) and relatively cheaper listing price?

Process

1. Loaded the data from Airbnb's New York City's public data (2008 to 2015)
2. Cleaned data (duplicates, outliers)
3. Explored data using visualizations (looked for trends)
4. Formed questions and investigate data
5. Created interactive dashboards in Tableau

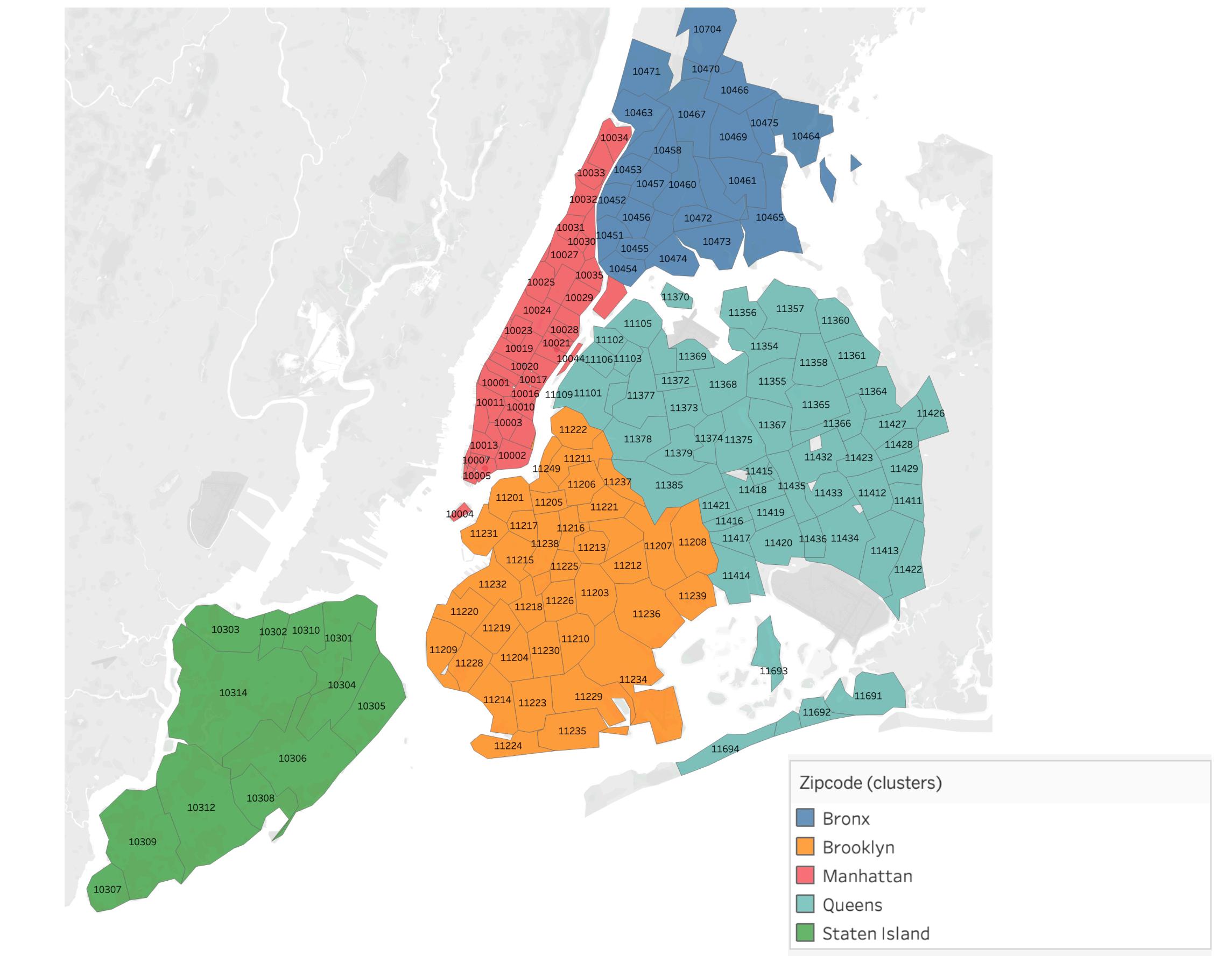
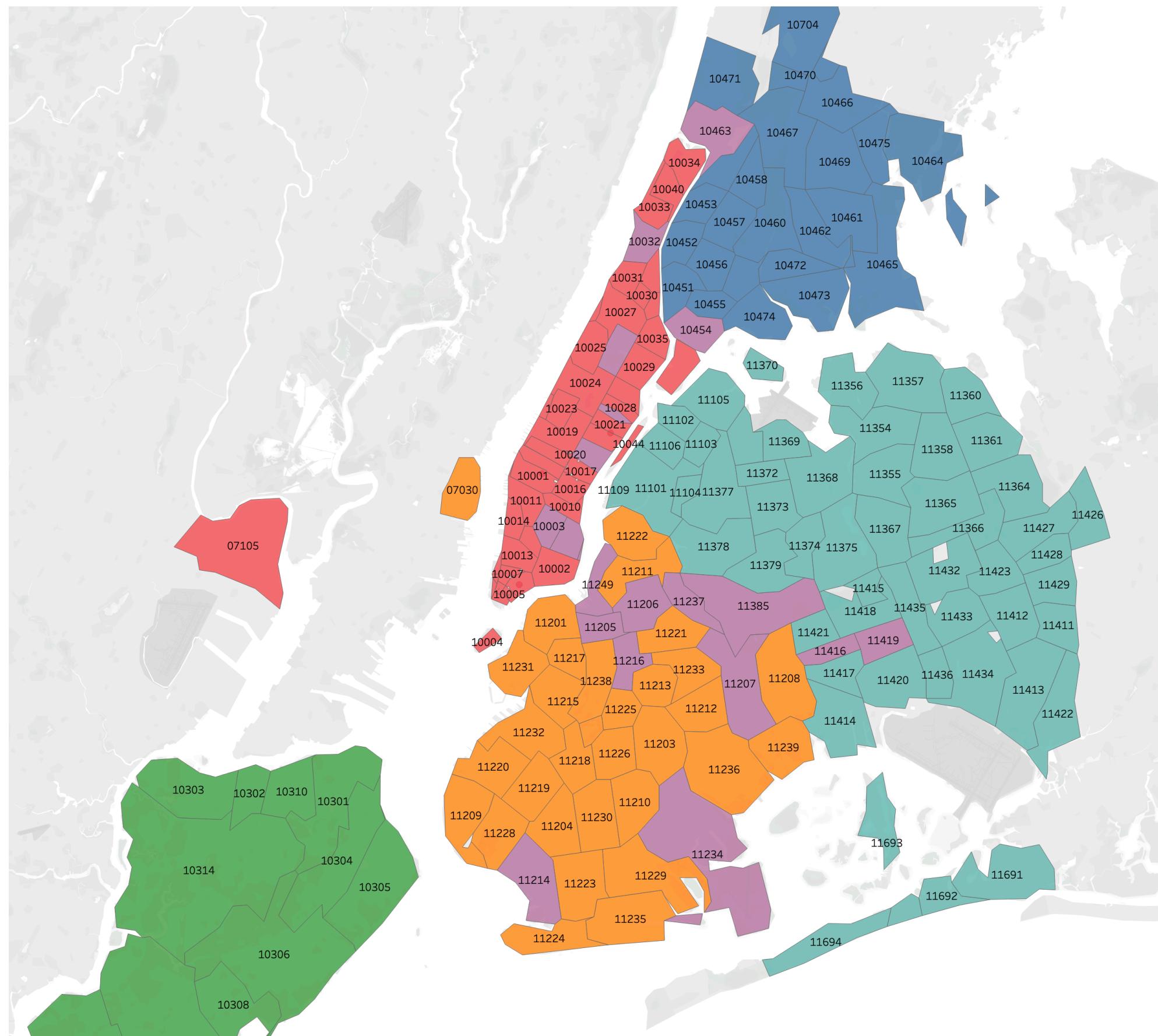
Data Field Exploration

Data Fields

Remote Field Name	Datatype	Definition
Host Id	Number	Airbnb's unique identifier for the host/user
Host Since	Date	The date the host/user was created. For hosts that are Airbnb guests this could be the date they registered as a guest.
Name	String	Name of the listing
Neighbourhood	String	Neighbourhood in NYC
Property Type	String	Self selected property type. Hotels and Bed and Breakfasts are described as such by their hosts in this field
Review Scores Rating (bin)	Number	
Room Type	String	All homes are grouped into the following three room types: Entire place / Private room / Shared room / Entire place
Zipcode	Geo	Geographic long-lat of listing
Beds	Number	The number of bed(s)
Number of Records	Number	The number of rows in the table
Number Of Reviews	Number	The number of reviews the listing has
Price	Number	daily price in local currency (USD)
Review Scores Rating	Number	Score rating from reviewers

Data Cleaning and Outlier Filtering

- Filtered out data with null zipcode value (133)
- Removed duplicates (4)
- Visually filtered data based on zipcode and neighbourhood mismatch
 - Fixed neighbourhood and zip code mismatches (using clustering)
 - 11237 -> Brooklyn
 - 11385 -> Queens
 - 10463 -> Bronx



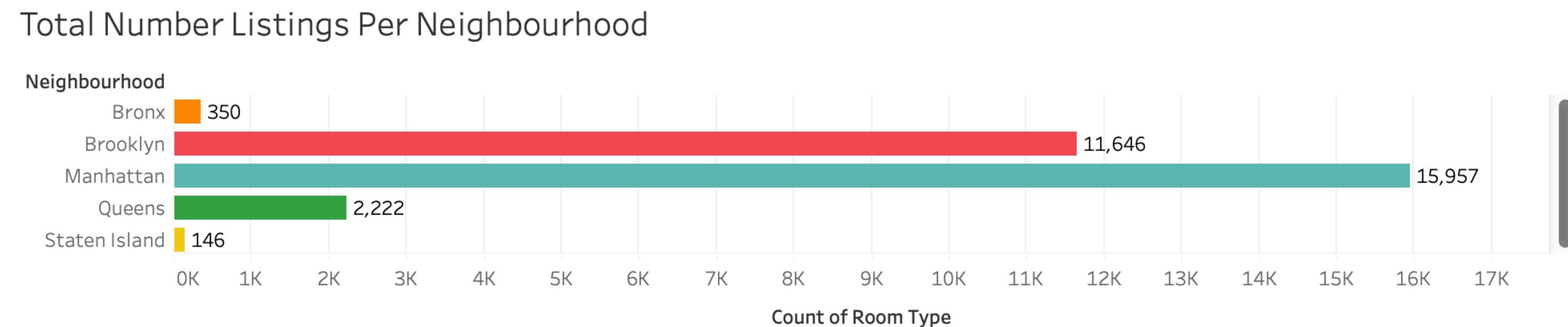
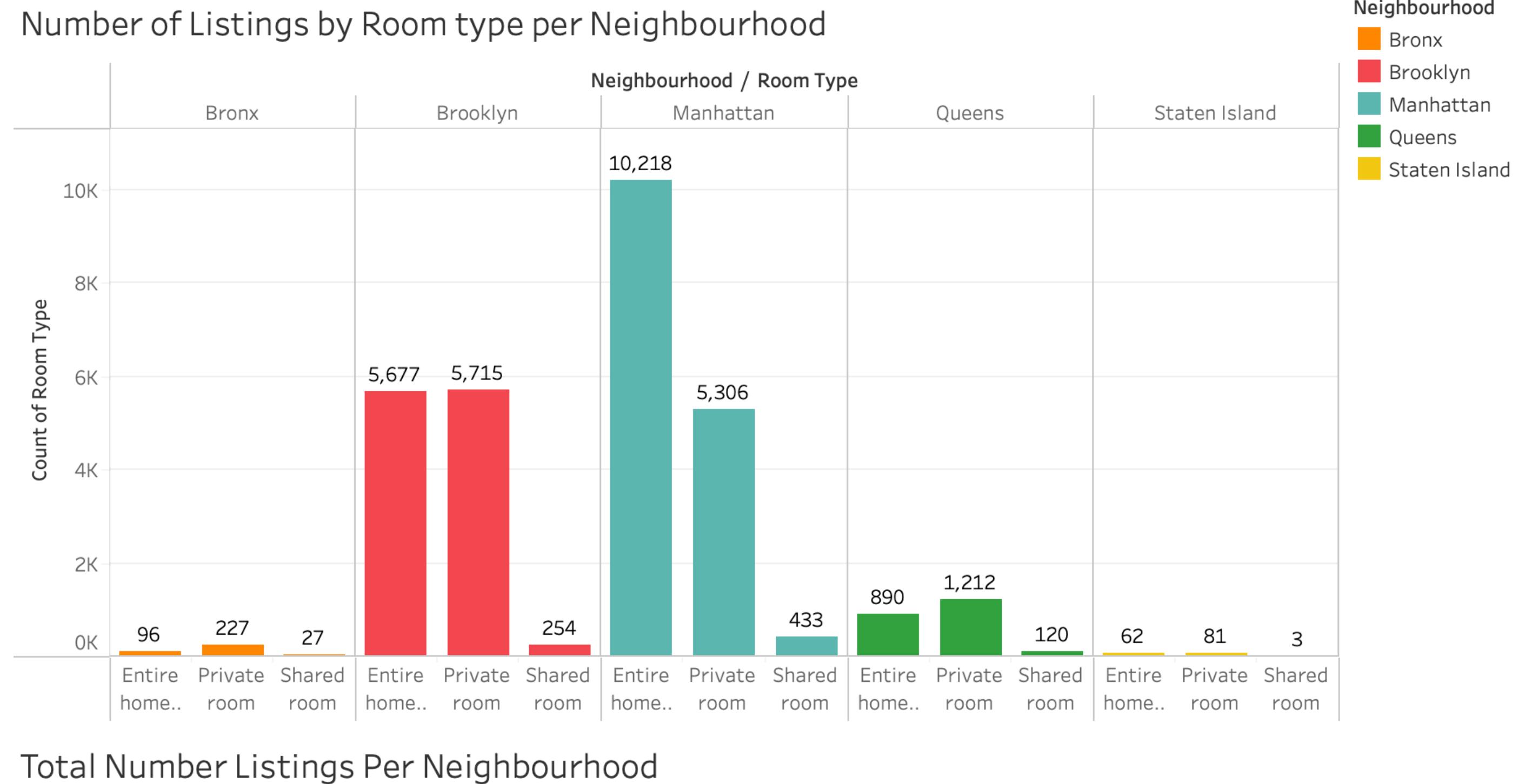
Data Cleaning and Outlier Filtering

Host ID	Neighbourhood	Zipcode	Name	Observation	Action
2334269	Manhattan	11216	Room in lovely Bed-Stuy brownstone!	Neighbourhood or zipcode wrong	Deleted
38483415	Manhattan	10454	Dreamer Room	Wrong neighbourhood	Changed neighbourhood to Bronx
3318518	Manhattan	11249	Brooklyn Loft with view of bridge	Wrong neighbourhood	Changed neighbourhood to Brooklyn
10234137	Manhattan	11206	Superbowl Studio Upper West Side	Wrong zipcode	Deleted
2321321	Queens	10003	Our Astorian Alcove	Wrong Zipcode	Deleted
3329066	Queens	11207	Pre-War charm 1 bdrm in Bushwick	Wrong neighbourhood based on name and zipcode	Changed neighbourhood to Brooklyn
21987837	Queens	11234	NYC Girls dorm in Astoria	Wrong neighbourhood	Deleted
10243387	Queens	11205	"Sweet Fort Green "	Wrong neighbourhood	Deleted
34919687	Queens	94103	Quarto Compartilhado em Astória-NY	Wrong zipcode, currently located in California	Deleted based on and null rating score and no reviews
14609026	Manhattan	99135	West Village 1bd sunnycool patio!!!	Wrong zipcode, currently located in Washington	Deleted based on and null rating score and no reviews
37083570	Manhattan	01003	Large historic loft style apartment	Wrong zipcode due to data	Corrected zipcode to Manhattan (10030)
7499240	Manhattan	07712	Beautiful Modern Asbury Apartment	Wrong zipcode, currently located in Monmouth County, New Jersey	Deleted based on and null rating score and no reviews
25912269	Manhattan	07105	Studio in Ironbound	Wrong neighbourhood, currently located in Manhattan	Deleted because Ironbound is in NJ
4105376	Brooklyn	07030	Large 1BR with a view - Brooklyn	Wrong zipcode, currently located in Hudson County, New Jersey	delete based null rating score and no reviews
1871955	Brooklyn	10022	Private Room in 2 Bdrm Apt. w/ Yard	Wrong zipcode, currently located in Manhattan but Neighbourhood is Brooklyn	delete based null rating score and no reviews
37034890	Brooklyn	10009	East Village Apartment	Wrong neighbourhood, currently Brooklyn	delete based null rating score and no reviews
7851685	Bronx	10032	historic townhouse on subway line	Wrong zipcode or neighbourhood	delete based on conflicting data
16958437	Manhattan	11419	BEAUTIFUL, HUGE STUDIO IN HARLEM!	Wrong zipcode	delete based null rating score and no reviews
6165150	Brooklyn	11416	CHARMING APARTMENT NEAR A/J TRAINSx	Wrong zipcode or neighbourhood	delete based on conflicting data
36710591	Queens	10026	Quaint room in 2bedroom apt	Wrong zipcode or neighbourhood	delete based on conflicting data
1410697	Queens	10075	Oversized 2 bedroom apartment	Wrong zipcode or neighbourhood	delete based on conflicting data

Data Exploration: Listings & Room Type per Neighbourhood

Summary:

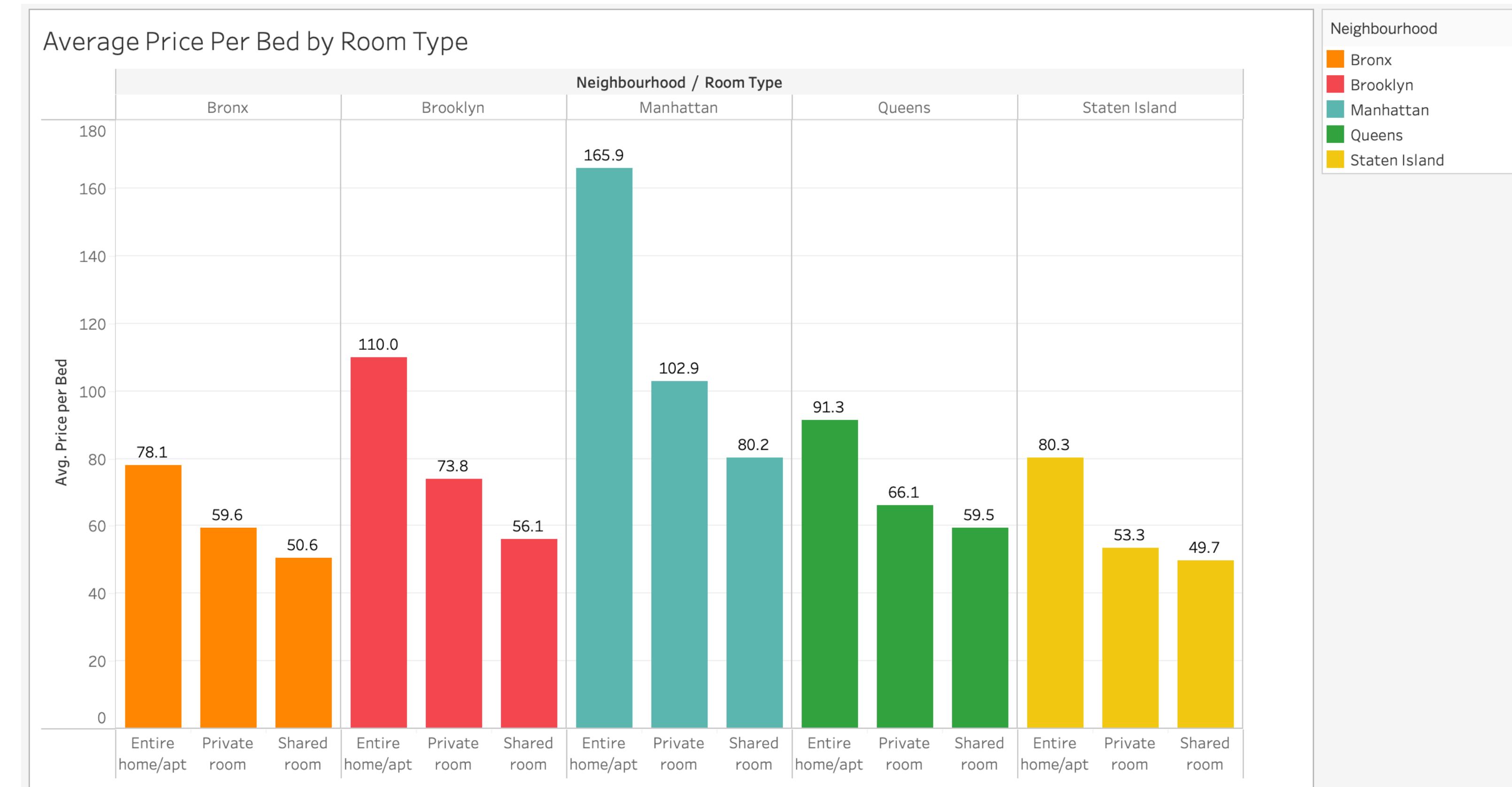
- Manhattan and Brooklyn have the most listings compared to the other neighbourhoods. This may be due to the high popularity of tourism in the area.
- Bronx, Queens, and Staten Island have more listings for private rooms than an entire home and shared room.



Data Exploration: Average Price per Bed

Summary:

- The average price of a bed increases with the space of room provided (i.e., the entire home/ apt)
- The average price of a bed for an entire home/ apartment is observed to be the most expensive in all neighbourhoods.
- Manhattan has the most expensive room types



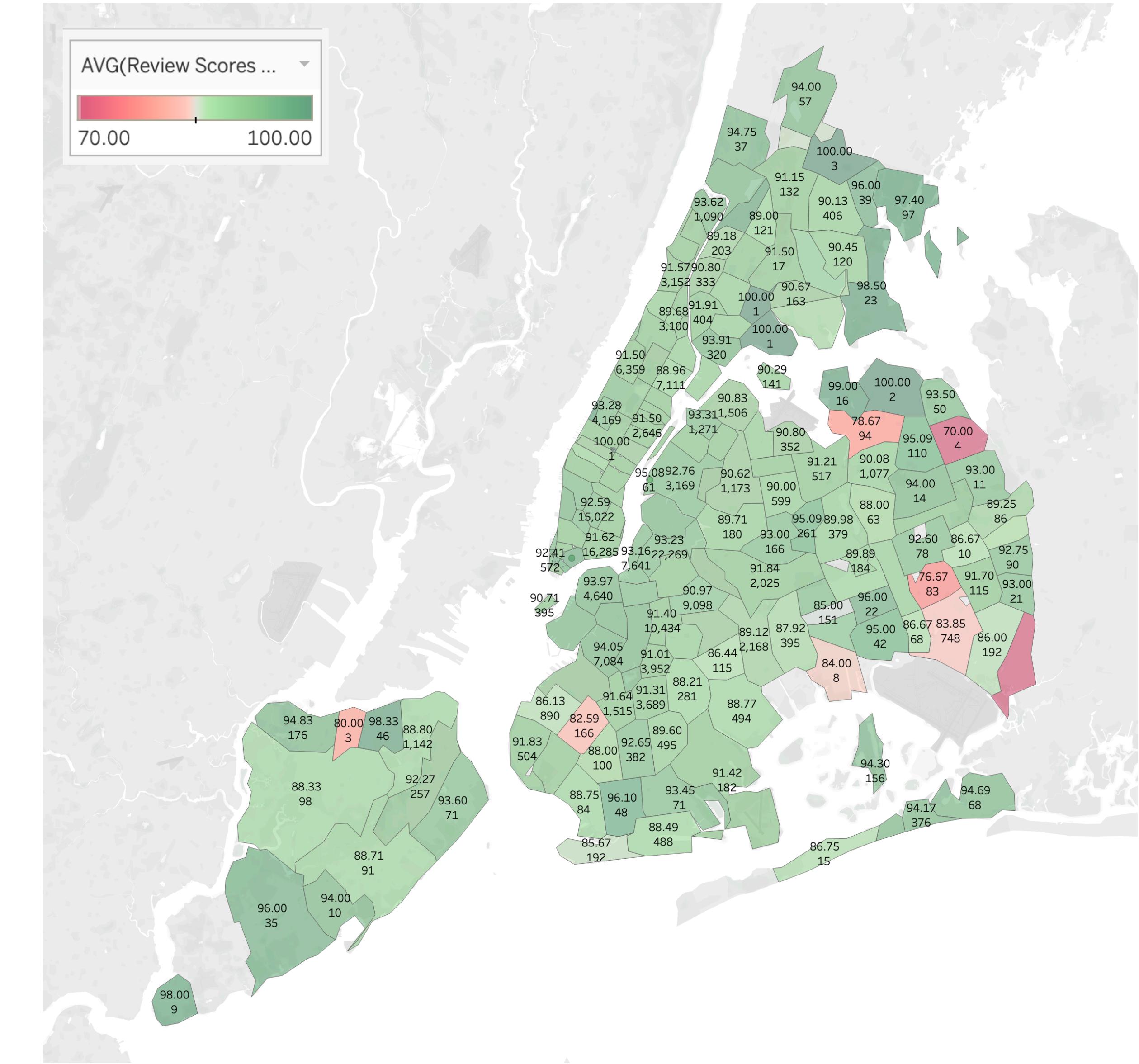
Data Exploration: Zipcode & Avg. Rating Score

Summary:

- The average rating score per zipcode are overall good with majority of zipcodes above 85% rating
- There are particular zipcodes with low rating scores that could be associated with a low number of ratings which can skew the average rating, the same is true with high rating listings

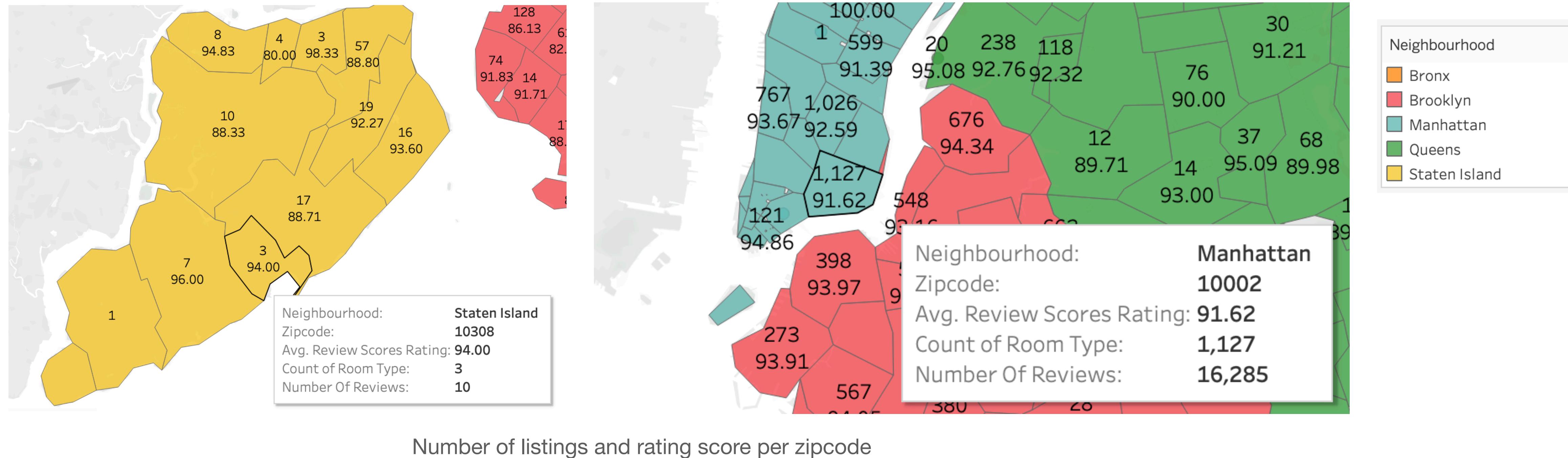
Zipcode	Neighbourhood	Avg Review Score	# of Reviews
10302	Staten Island	80.00	3
11361	Queens	70.00	4
11414	Queens	84.00	8
11422	Queens	70.00	4

Low Avg. Rating Score Zipcodes



Data Exploration: Zipcode & Avg. Rating Score (cont'd)

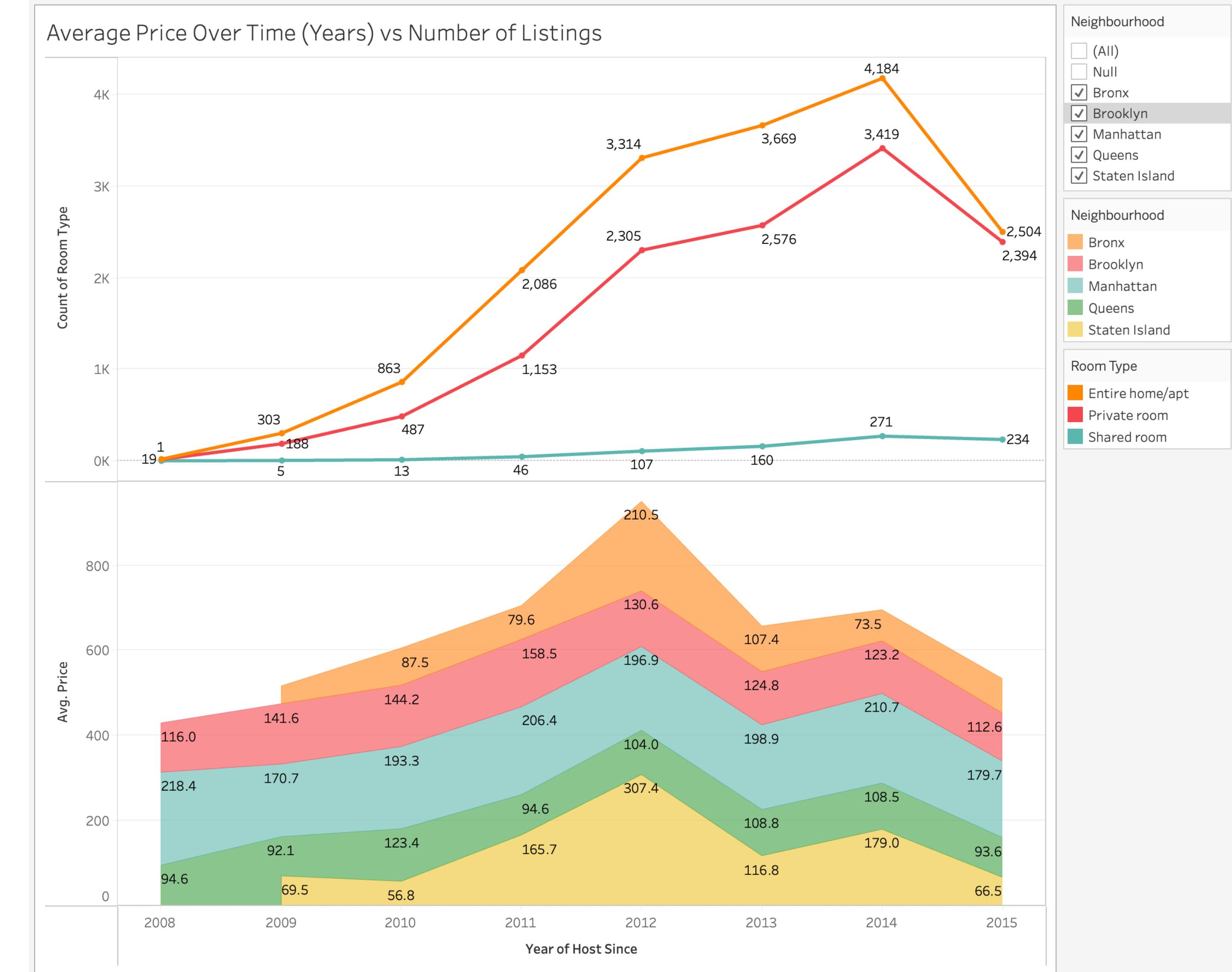
- There are particular zipcodes with low rating scores that could be associated with a low number of reviews which can skew the average rating (the same is true with high rating listings)
- For example, in Staten Island for the zipcode 10308, there were only 3 listings (count of room type), but the average rating score is very high (94).



Data Exploration: Avg. Price vs. Number of Listings

Summary:

- There is an average price of listings increase for every NYC neighbourhood until 2012. After 2012, there is a sharp decline in average price for each neighbourhood per year.
- The sharp decline in the average price for listings may be due to the increase of listings in 2012 (based on the number of room types)
- The average price of listings declines tapers off between 2013 and 2014. This could be due to the continual and gradual increase of listings.
- Between 2014 and 2015, there is a decline in average price of listings. This could be due to a mixture of competition (increase of listings) and increased regulations for short-term housing rentals in NYC. Airbnb removed 1500 listings from the public data released from November 2014 to November 2015.



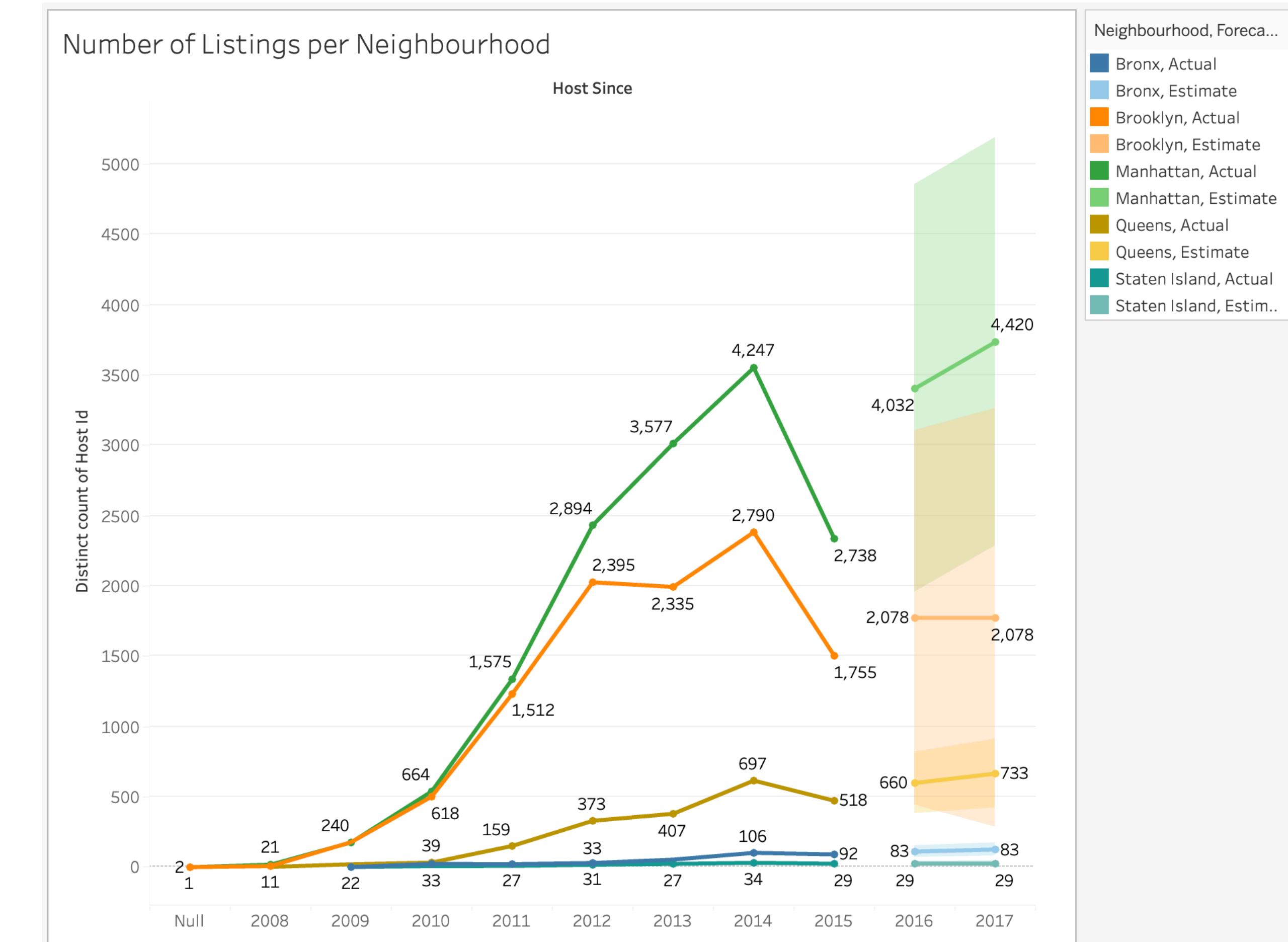
Forecast Listing by Neighbourhood

Summary:

- The number of listed hosts have increased from 2008 to 2015
- Manhattan and Brooklyn saw the most number of increases in listings compared to Queens, Staten Island and the Bronx
- With forecast analysis, Manhattan and Queens are both projected to increase number of listings by ~10% from 2016-2017

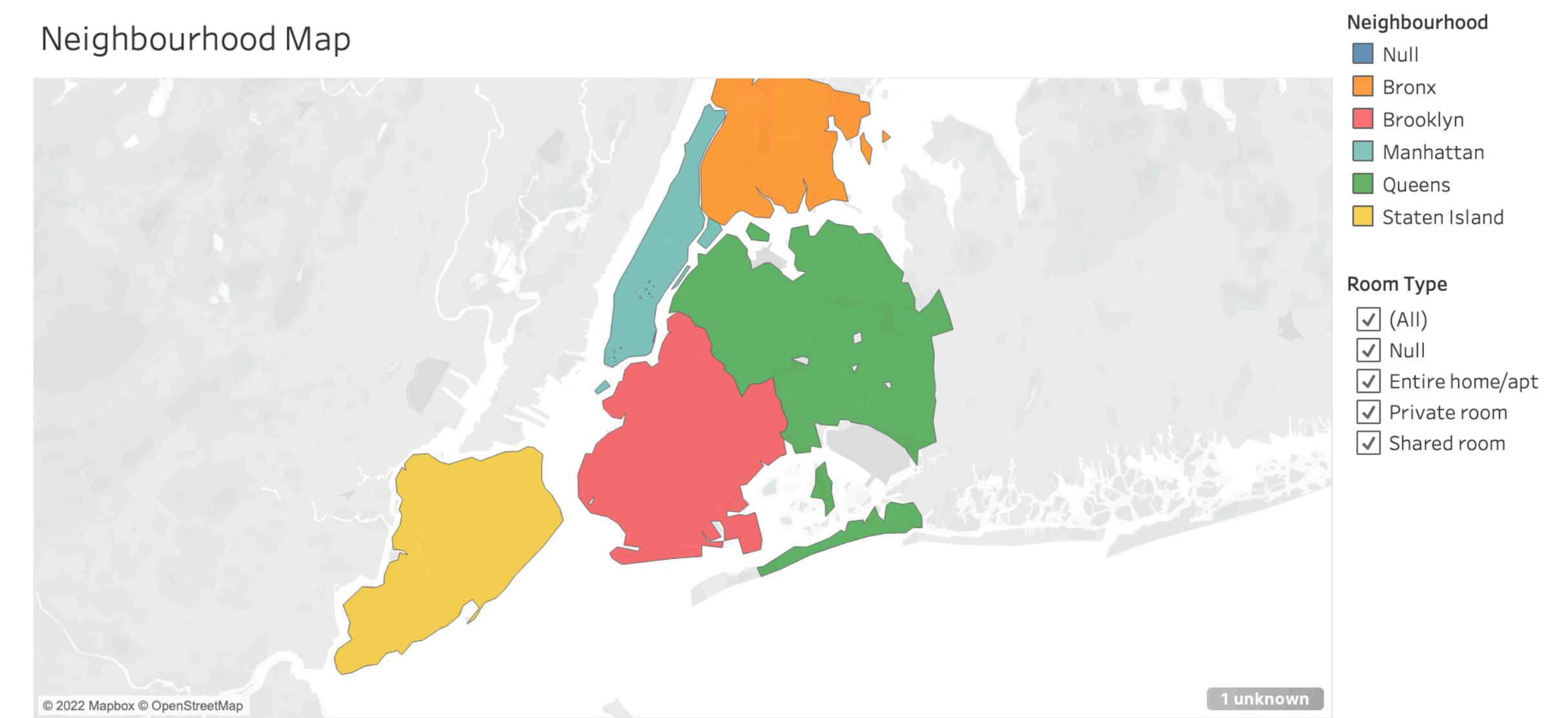
Interesting Finding:

- Airbnb removed ~2000 listings in 2015 after NY attorney general reports 72% of Airbnb listings violate either New York State or New York City regulations
- <https://www.bloomberg.com/news/articles/2014-10-16/new-york-s-attorney-general-declares-war-on-airbnb>
- In 2015, Airbnb released its data from November 2014 to November 2015 to the public, but Airbnb removed 1500 listings from the data, which makes one question the validity of the released data
- <https://skift.com/2015/12/01/airbnbs-data-on-new-york-city-business-shows-most-hosts-break-the-law/>

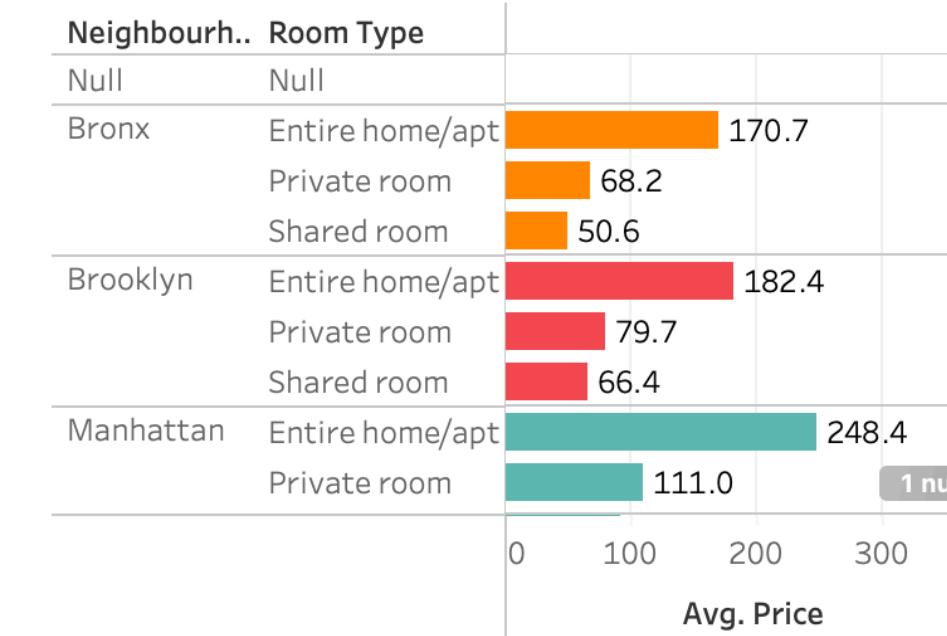


Question 1

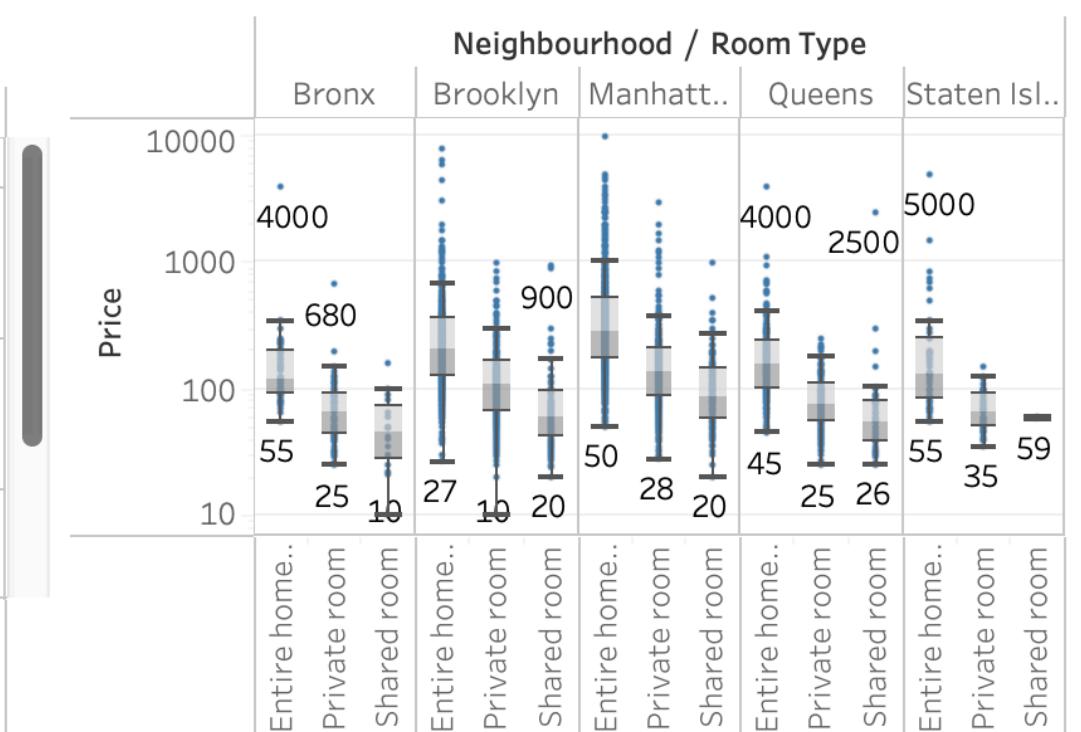
What is the median price for each room type in a particular neighbourhood?



Average Price by Room Type per Neighbourhood

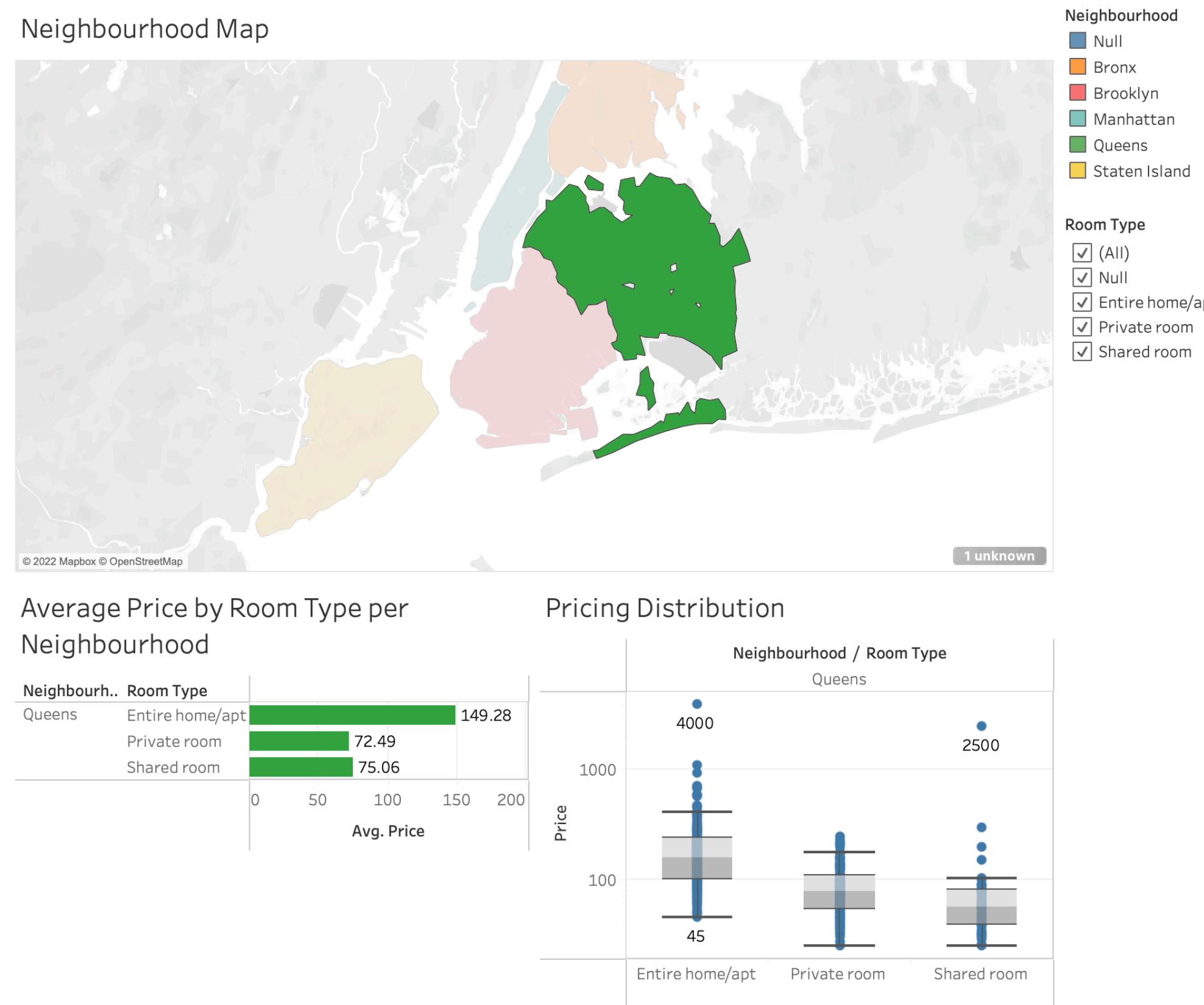


Pricing Distribution



Question 1

What is the median price for each room type in a particular neighbourhood?

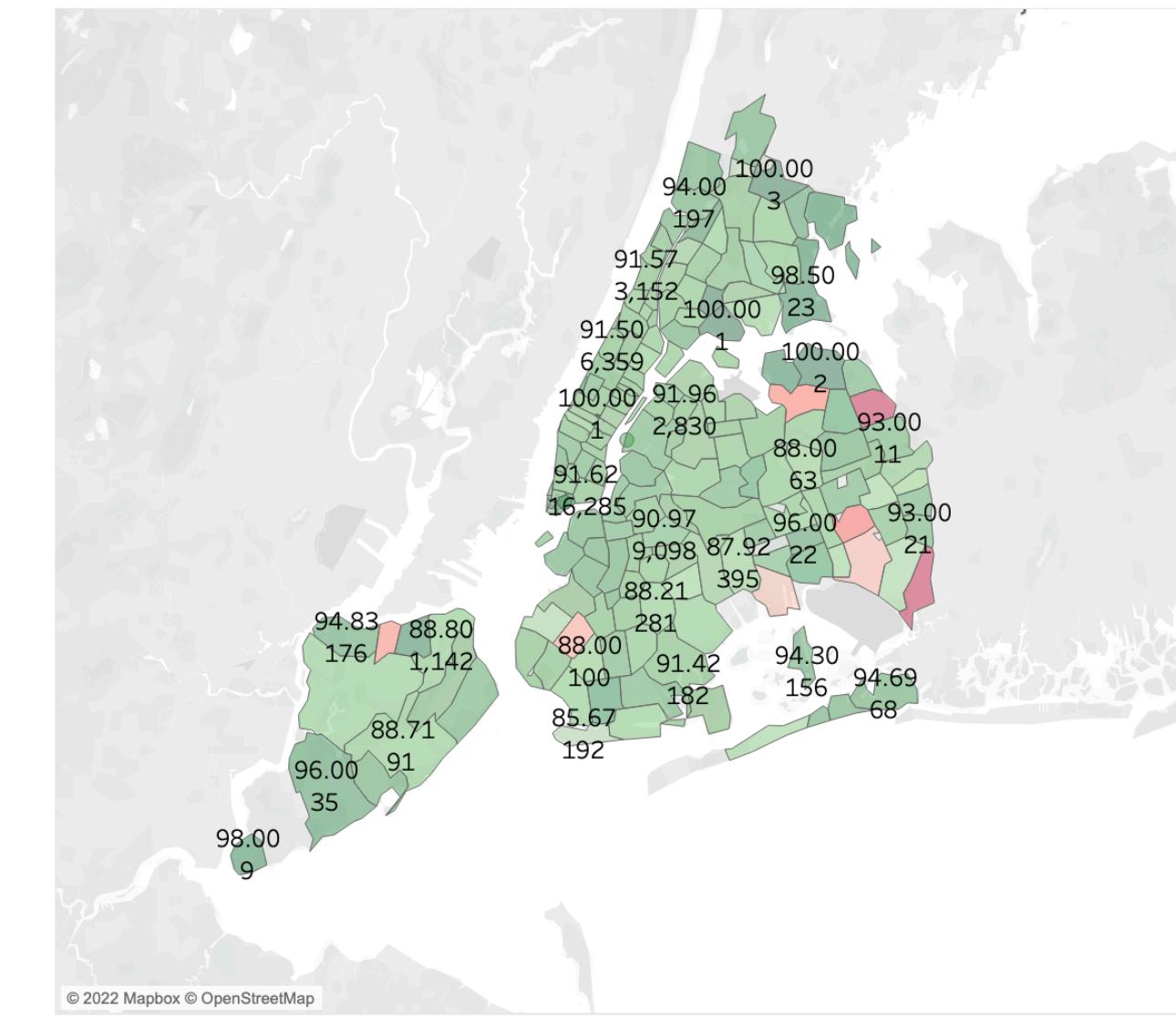


For example, in Queens, the average price for entire home/apt is \$149.28. However, this number may have been skewed by the outlier of a \$4000 listing as shown in the price distribution.

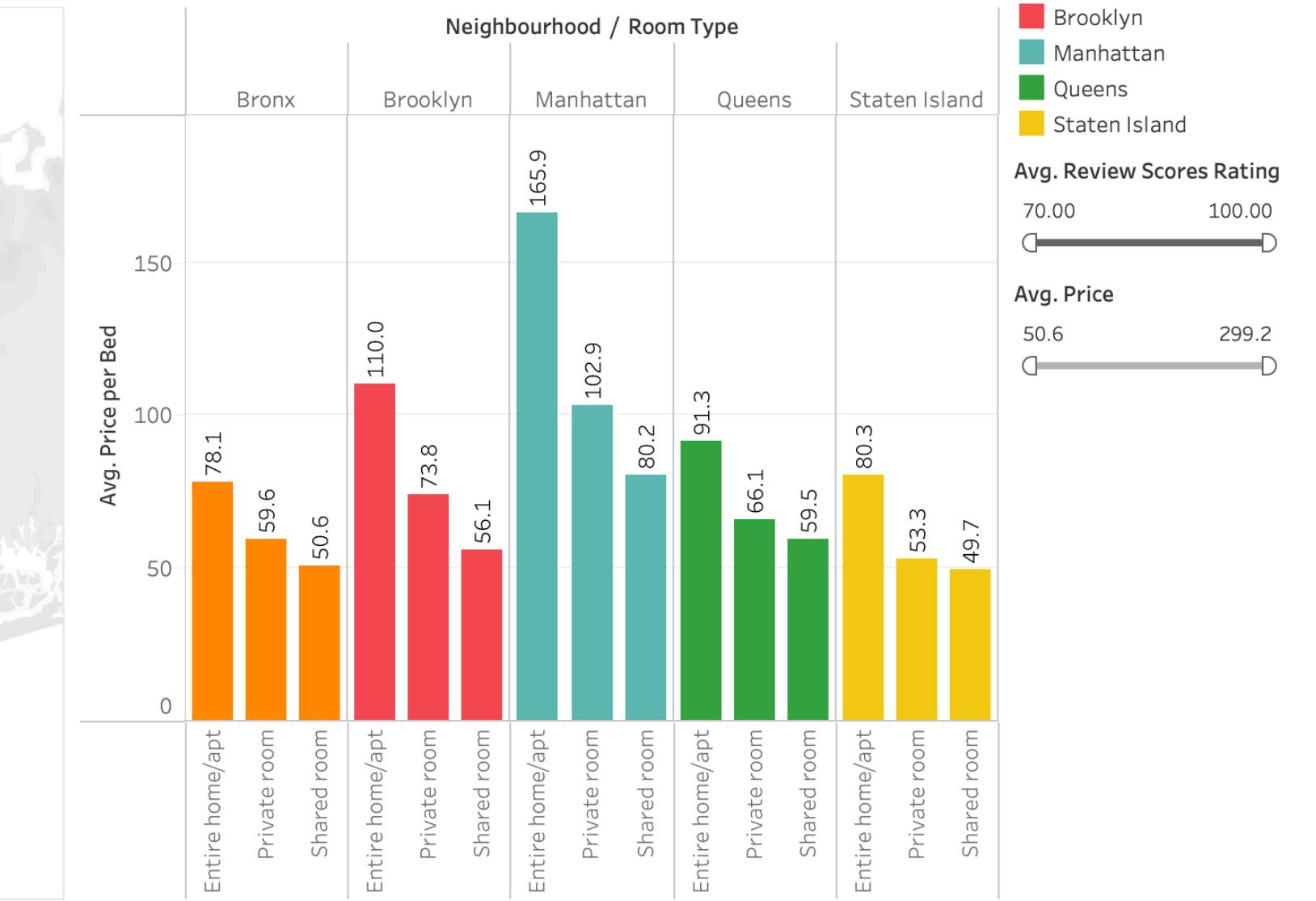
Question 2

What are the room types in a particular zipcode and their respective prices and rating score?

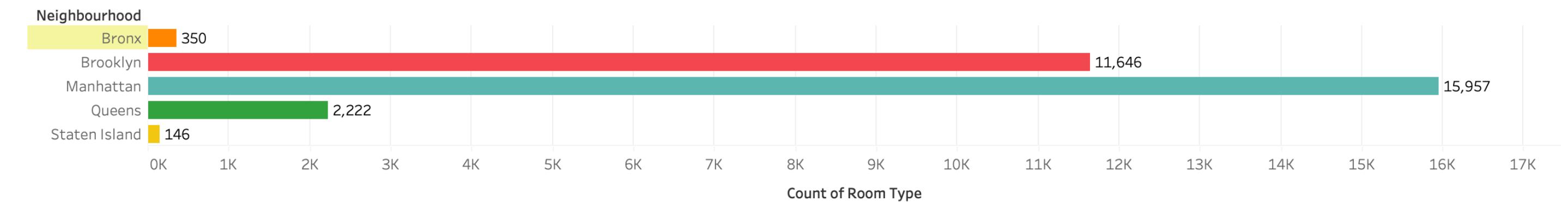
Average Rating Score by Zipcode



Average Price Per Bed by Room Type

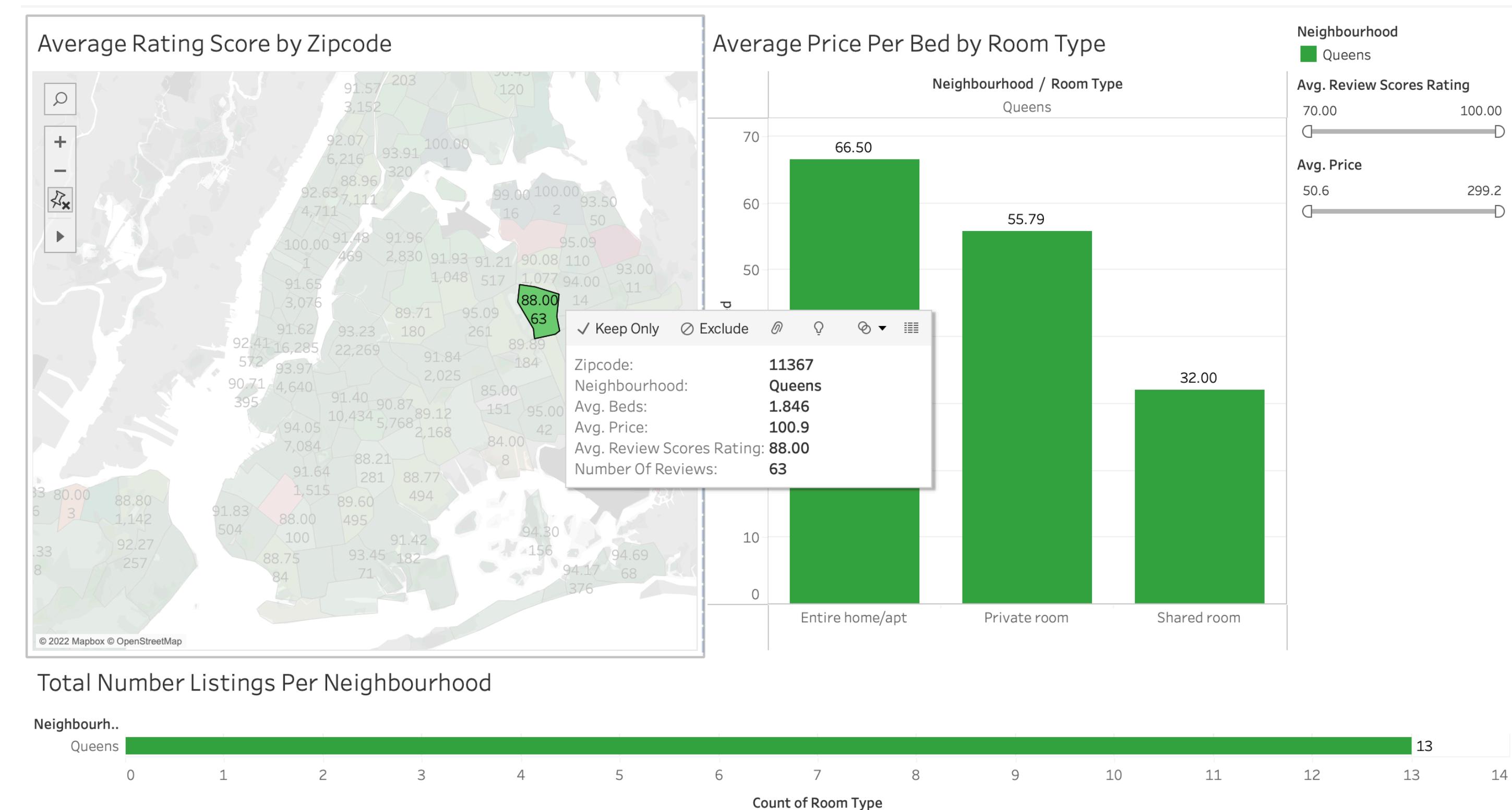


Total Number Listings Per Neighbourhood



Question 2

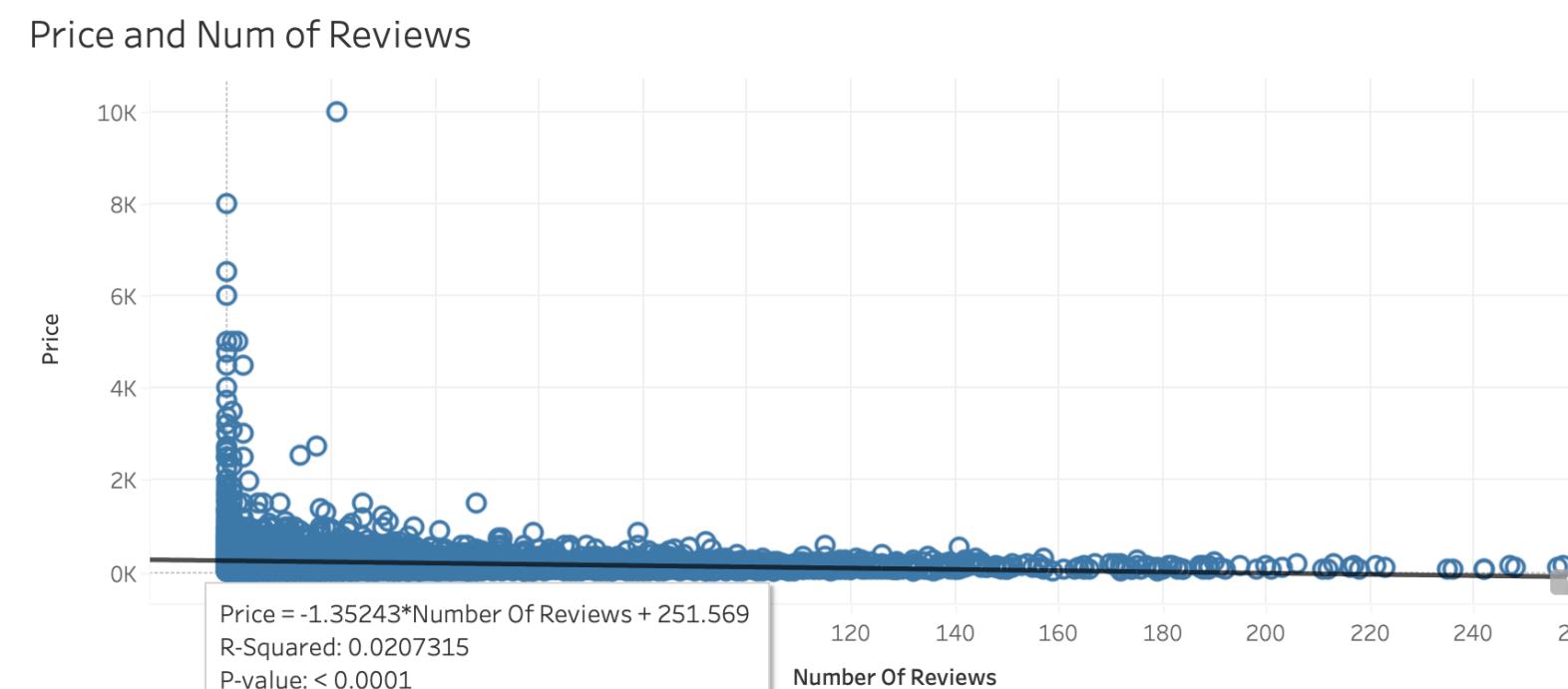
What are the room types in a particular zipcode and their respective price per bed and rating score?



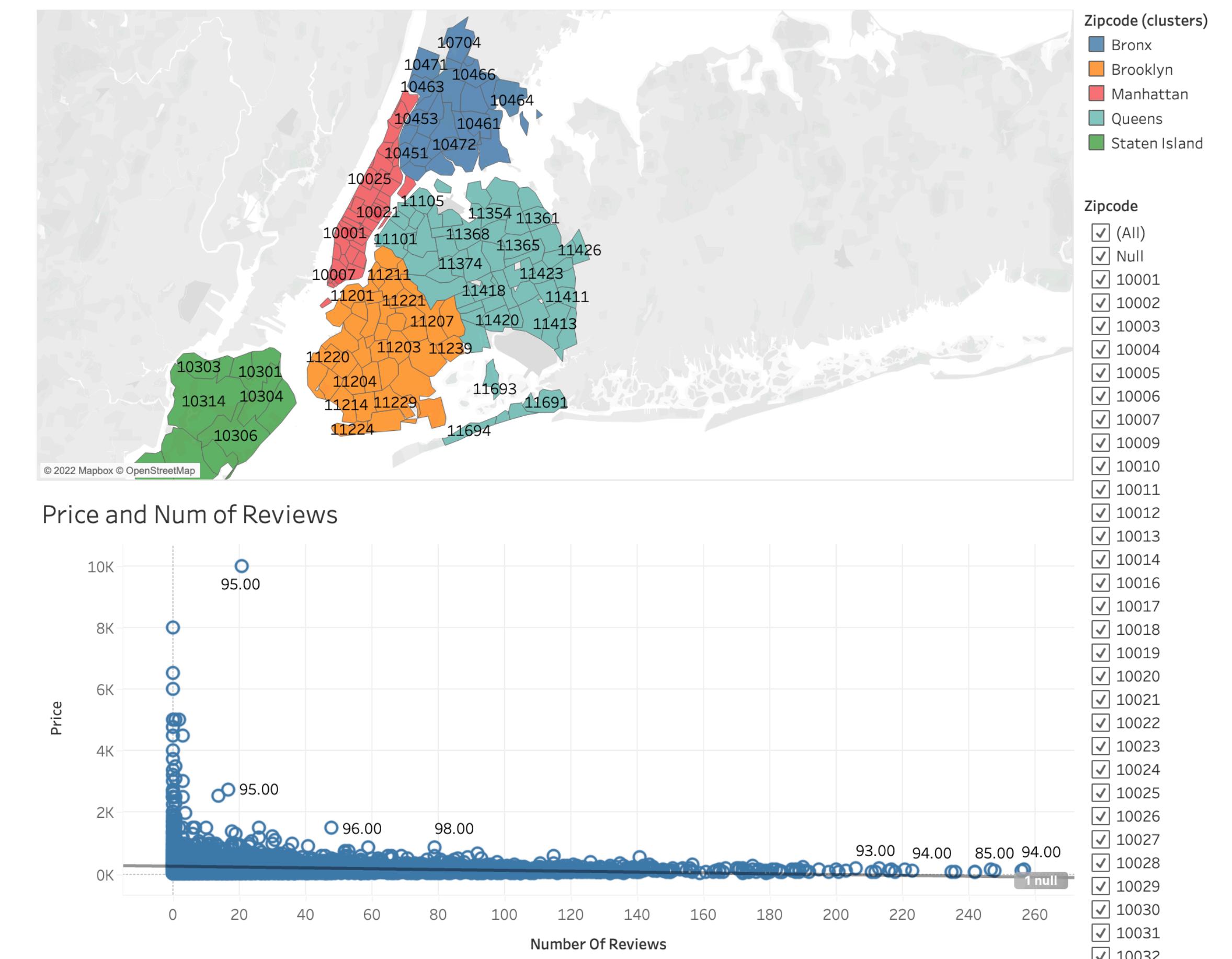
For example, in a particular zipcode (area) in Queens, where the average review score rating is 88 with 63 number of reviews, there are 13 listings and the average price per bed for an *Entire home/apt* is \$66.50.

Question 3

Which zipcodes have the highest ratings (with over 100 number of reviews) and relatively cheaper listing price?

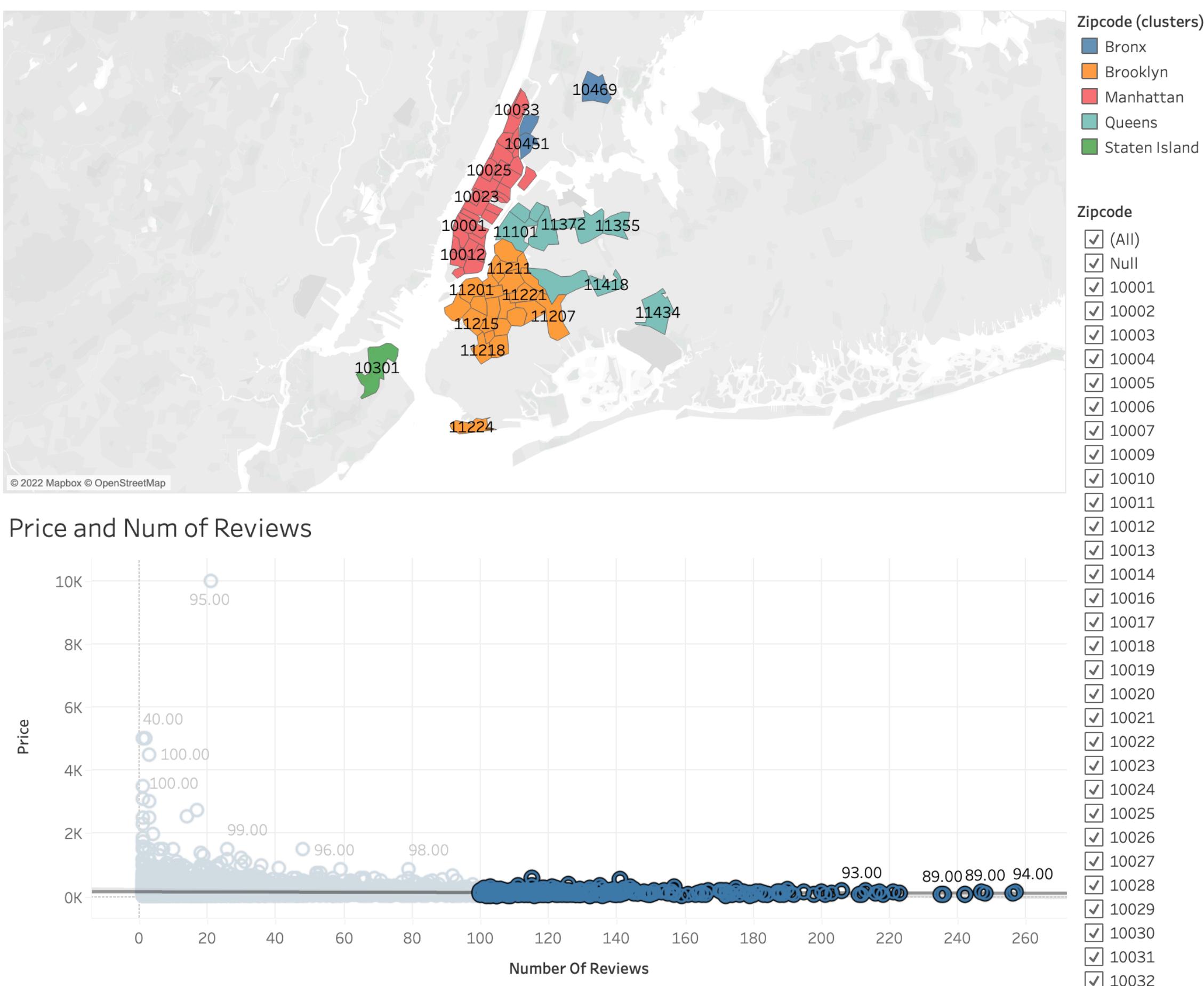


There is no linear relationship between the price of the listings and the number of reviews (R-squared 0.0207315). However, the p-value proves our linear model to be significant (<0.0001).



Question 3

Which zipcodes have the highest ratings (with over 100 number of reviews) and relatively cheaper listing price?



Challenges

1. Time constraint to clean data and filter outliers

- Mistake: Missed cleaning up the column names as there were duplicate listings (Name) for unique Host ID.
- Future Action: Further investigate and remove or compile duplicates fields.

Host Id	Host Since	Name	Neighbourhood	Property Type	Review Scores Rating (bin)	Room Type	Zipcode	Beds	Number Of Reviews	Price	Review Scores Rating
417504	2011-03-01	Loft Suite @ Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	16	199	95
417504	2011-03-01	Loft Suite @ Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	14	199	97
417504	2011-03-01	Loft Suite @ Box House Hotel	Brooklyn	Apartment	100	Entire home/apt	11222	2	5	199	100
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	85	Entire home/apt	11222	2	7	199	88
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	90	Entire home/apt	11222	2	6	199	92
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	90	Entire home/apt	11222	2	5	249	93
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	6	199	97
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	8	199	97
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	14	199	98
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	17	199	98
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	18	199	98
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	20	199	98
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	100	Entire home/apt	11222	2	2	199	100
417504	2011-03-01	Loft Suite @ The Box House Hotel	Brooklyn	Apartment	100	Entire home/apt	11222	2	6	199	100
417504	2011-03-01	Loft w/ Terrace @ Box House Hotel	Brooklyn	Apartment	80	Entire home/apt	11222	2	2	249	80
417504	2011-03-01	Loft w/ Terrace @ Box House Hotel	Brooklyn	Apartment	85	Entire home/apt	11222	2	5	349	88
417504	2011-03-01	Loft w/ Terrace @ Box House Hotel	Brooklyn	Apartment	95	Entire home/apt	11222	2	22	249	99
417504	2011-03-01	Loft w/ Terrace @ Box House Hotel	Brooklyn	Apartment	100	Entire home/apt	11222	2	2	249	100
417504	2011-03-01	Loft Suite @ Box House Hotel	Brooklyn	Loft	90	Entire home/apt	11222	2	3	199	93

2. Create better forecast for prices per listing

3. Download updated data

4. Have more fields to work with

Future Goals

- Get more data to have a more thorough investigation and accurate findings, such as investigating on the effects of crime rates on the price of listings
- Investigate which time of the year have more check-ins and forecast the popularity of occupancy