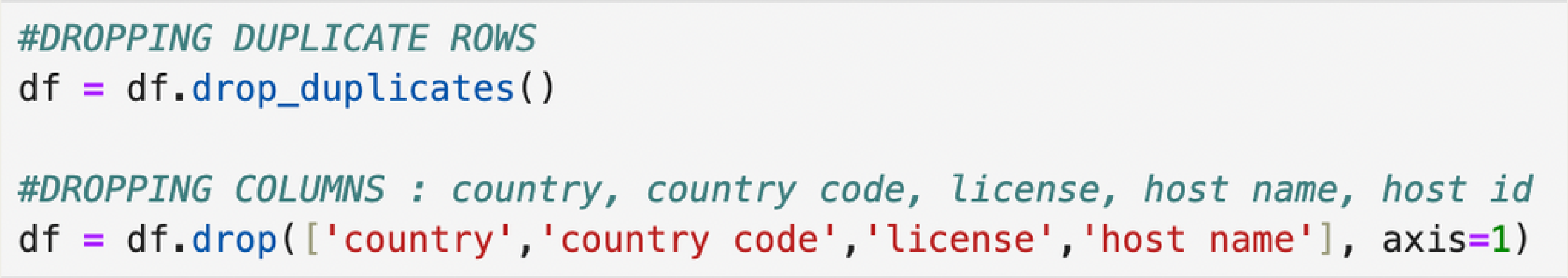
MSDS 593 Final Project

AIRBNB OPEN DATA

Bhumika Srinivas
Eren Bardak
Sissi Shen

INSIGHTFUL DATA EXPLORATION

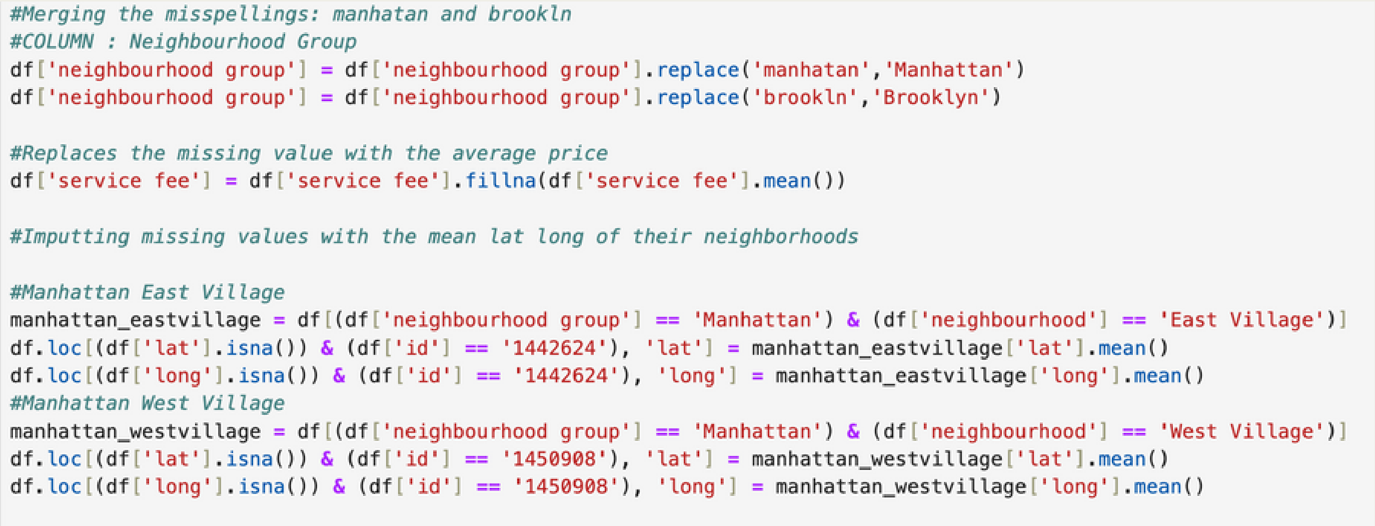
- Dropping duplicate rows and unnecessary columns



```
#DROPPING DUPLICATE ROWS
df = df.drop_duplicates()

#DROPPING COLUMNS : country, country code, license, host name, host id
df = df.drop(['country','country code','license','host name'], axis=1)
```

- Corrected misspellings and Imputed missing values



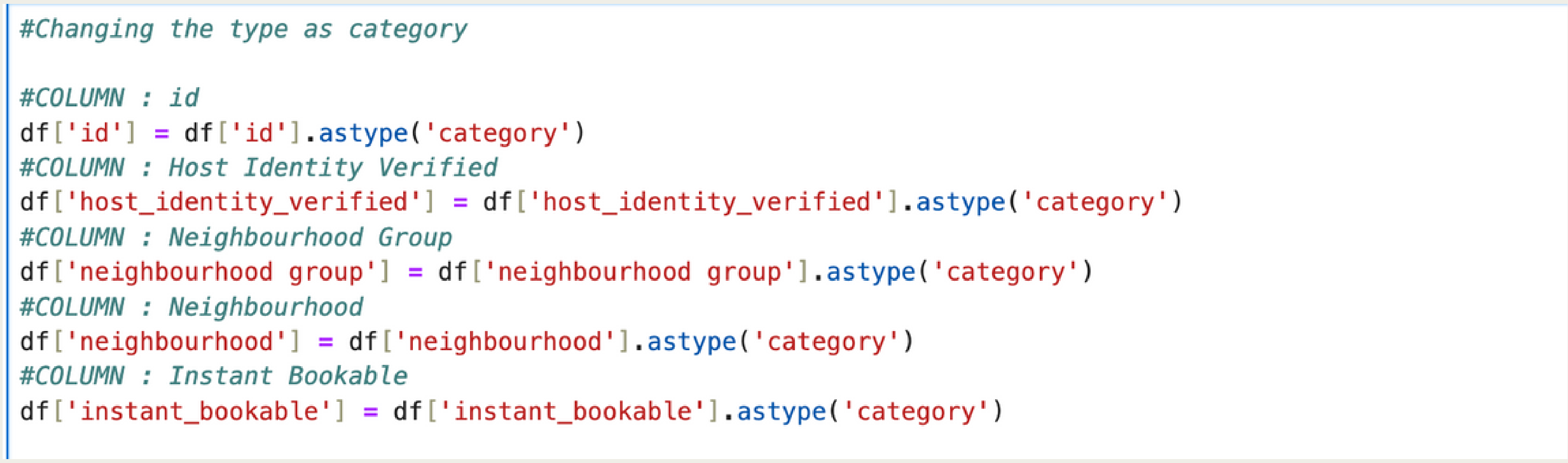
```
#Merging the misspellings: manhatan and brookln
#COLUMN : Neighbourhood Group
df['neighbourhood group'] = df['neighbourhood group'].replace('manhatan','Manhattan')
df['neighbourhood group'] = df['neighbourhood group'].replace('brookln','Brooklyn')

#Replaces the missing value with the average price
df['service fee'] = df['service fee'].fillna(df['service fee'].mean())

#Imputting missing values with the mean lat long of their neighborhoods

#Manhattan East Village
manhattan_eastvillage = df[(df['neighbourhood group'] == 'Manhattan') & (df['neighbourhood'] == 'East Village')]
df.loc[(df['lat'].isna()) & (df['id'] == '1442624'), 'lat'] = manhattan_eastvillage['lat'].mean()
df.loc[(df['long'].isna()) & (df['id'] == '1442624'), 'long'] = manhattan_eastvillage['long'].mean()
#Manhattan West Village
manhattan_westvillage = df[(df['neighbourhood group'] == 'Manhattan') & (df['neighbourhood'] == 'West Village')]
df.loc[(df['lat'].isna()) & (df['id'] == '1450908'), 'lat'] = manhattan_westvillage['lat'].mean()
df.loc[(df['long'].isna()) & (df['id'] == '1450908'), 'long'] = manhattan_westvillage['long'].mean()
```

- Change of data types

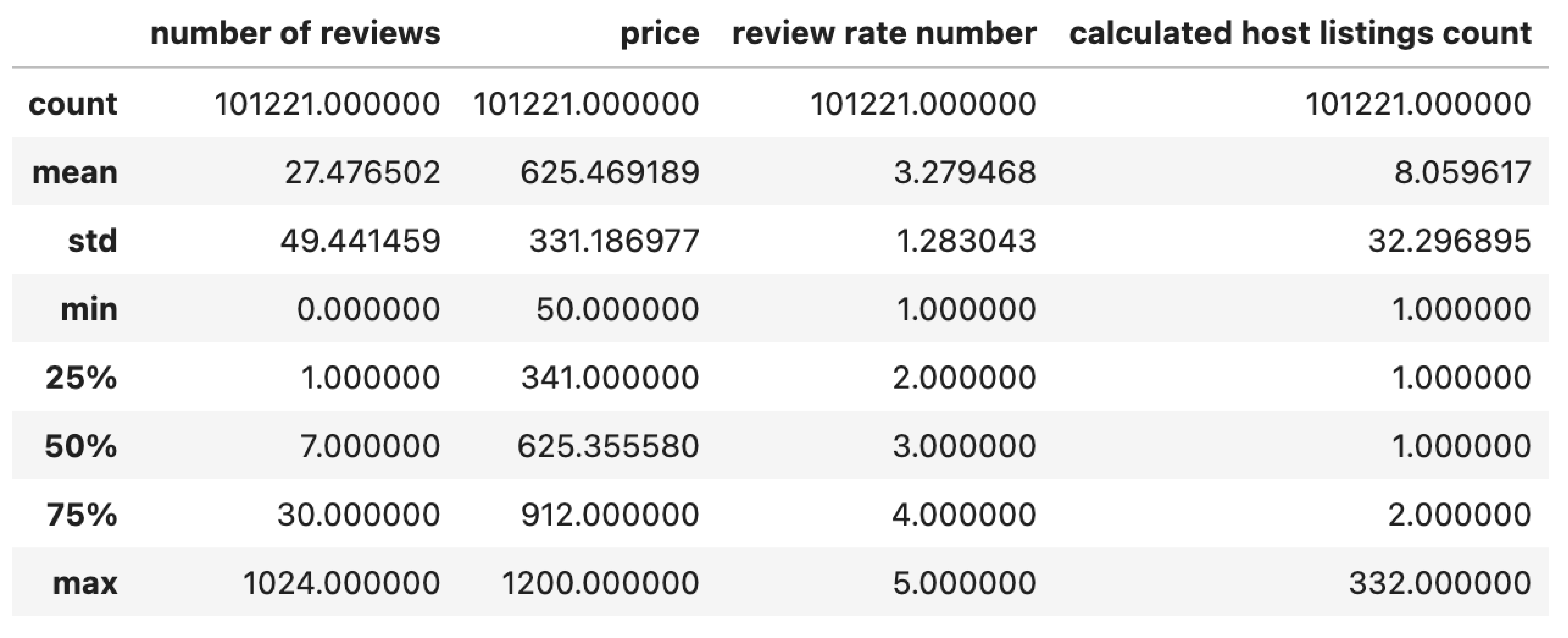


```
#Changing the type as category

#COLUMN : id
df['id'] = df['id'].astype('category')
#COLUMN : Host Identity Verified
df['host_identity_verified'] = df['host_identity_verified'].astype('category')
#COLUMN : Neighbourhood Group
df['neighbourhood group'] = df['neighbourhood group'].astype('category')
#COLUMN : Neighbourhood
df['neighbourhood'] = df['neighbourhood'].astype('category')
#COLUMN : Instant Bookable
df['instant_bookable'] = df['instant_bookable'].astype('category')
```

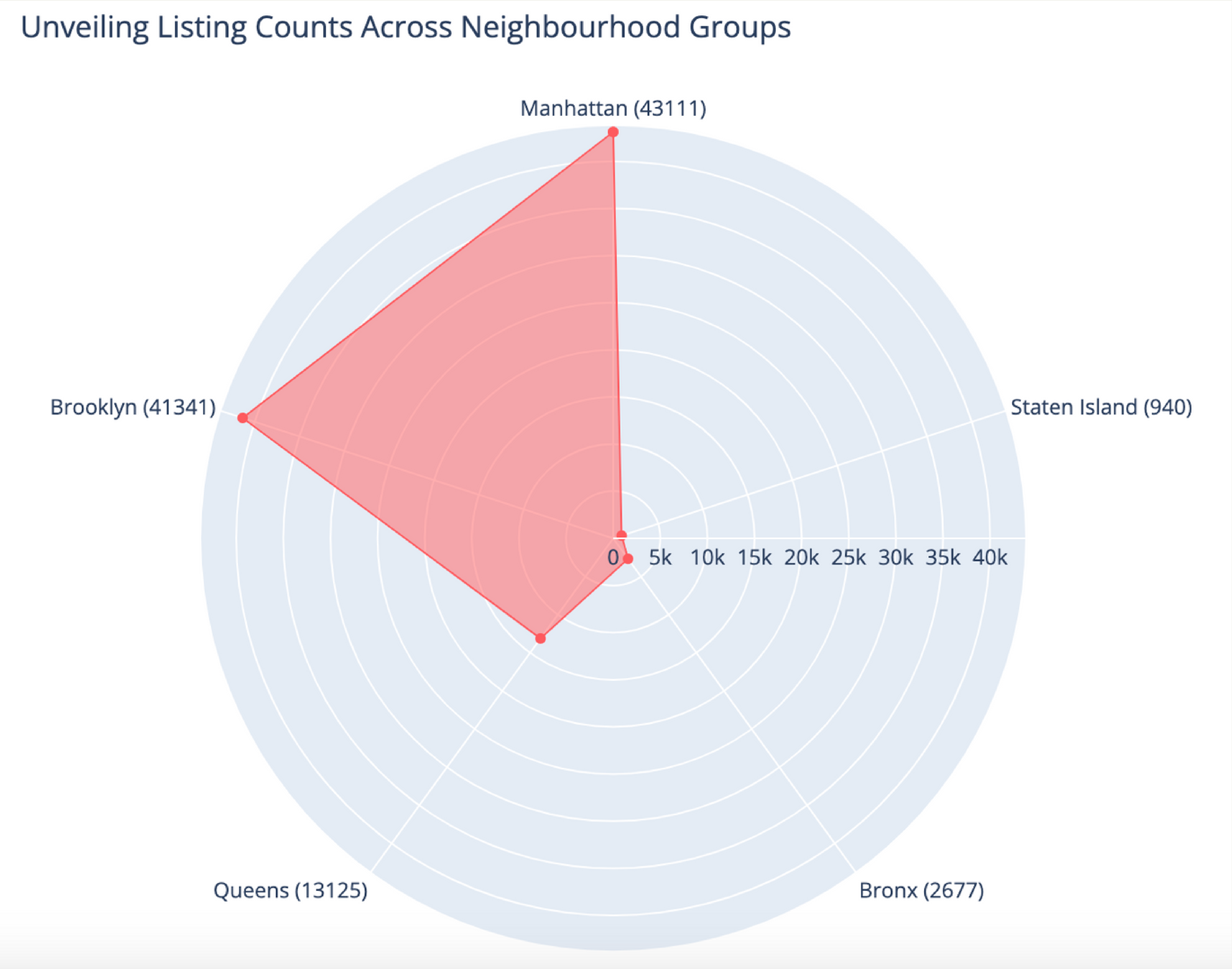
STATISTICAL SNAPSHOTS

26 Variables | 102599 Records → 15 Variables | 101221 Records

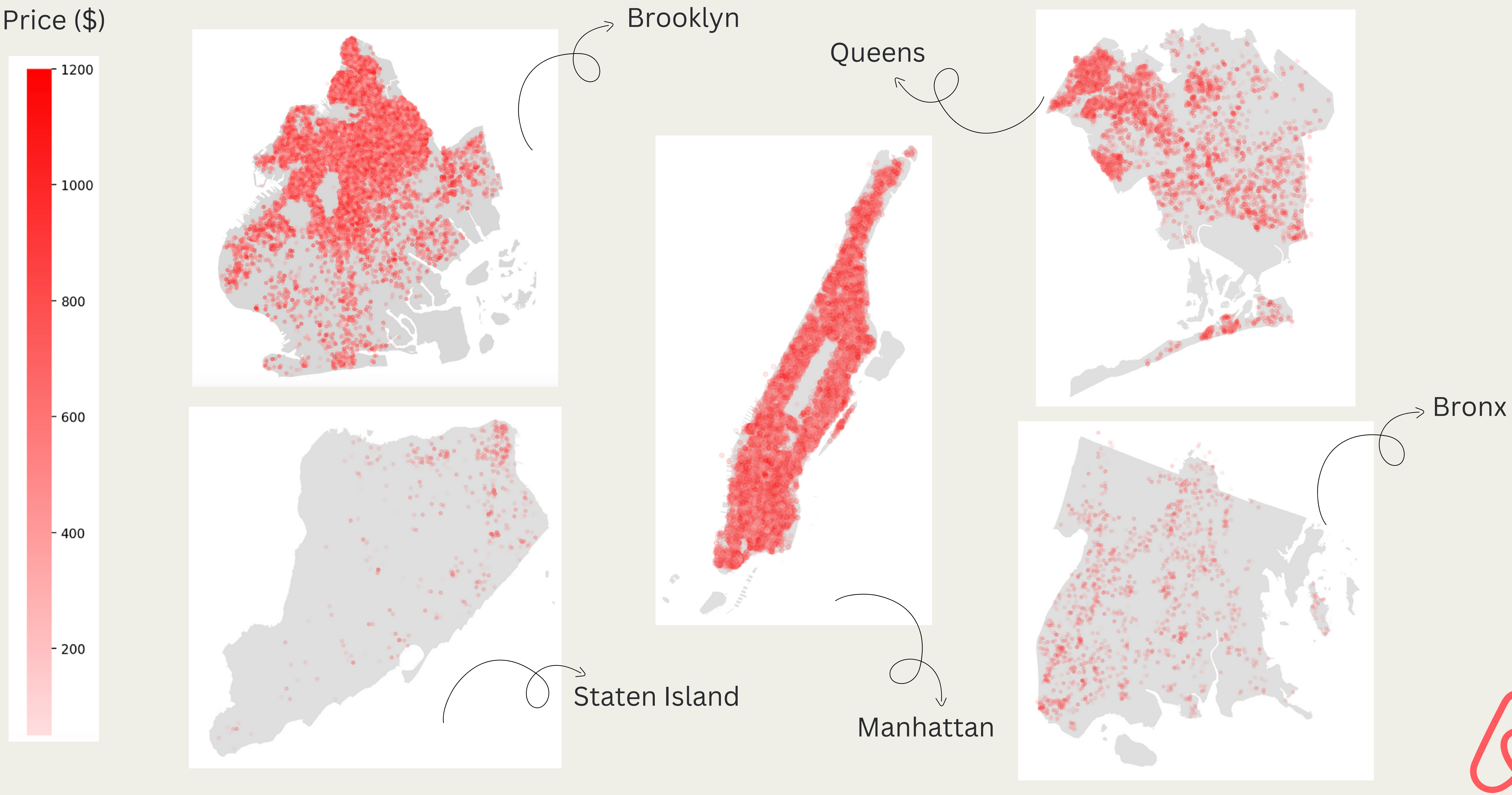


	number of reviews	price	review rate number	calculated host listings count
count	101221.000000	101221.000000	101221.000000	101221.000000
mean	27.476502	625.469189	3.279468	8.059617
std	49.441459	331.186977	1.283043	32.296895
min	0.000000	50.000000	1.000000	1.000000
25%	1.000000	341.000000	2.000000	1.000000
50%	7.000000	625.355580	3.000000	1.000000
75%	30.000000	912.000000	4.000000	2.000000
max	1024.000000	1200.000000	5.000000	332.000000

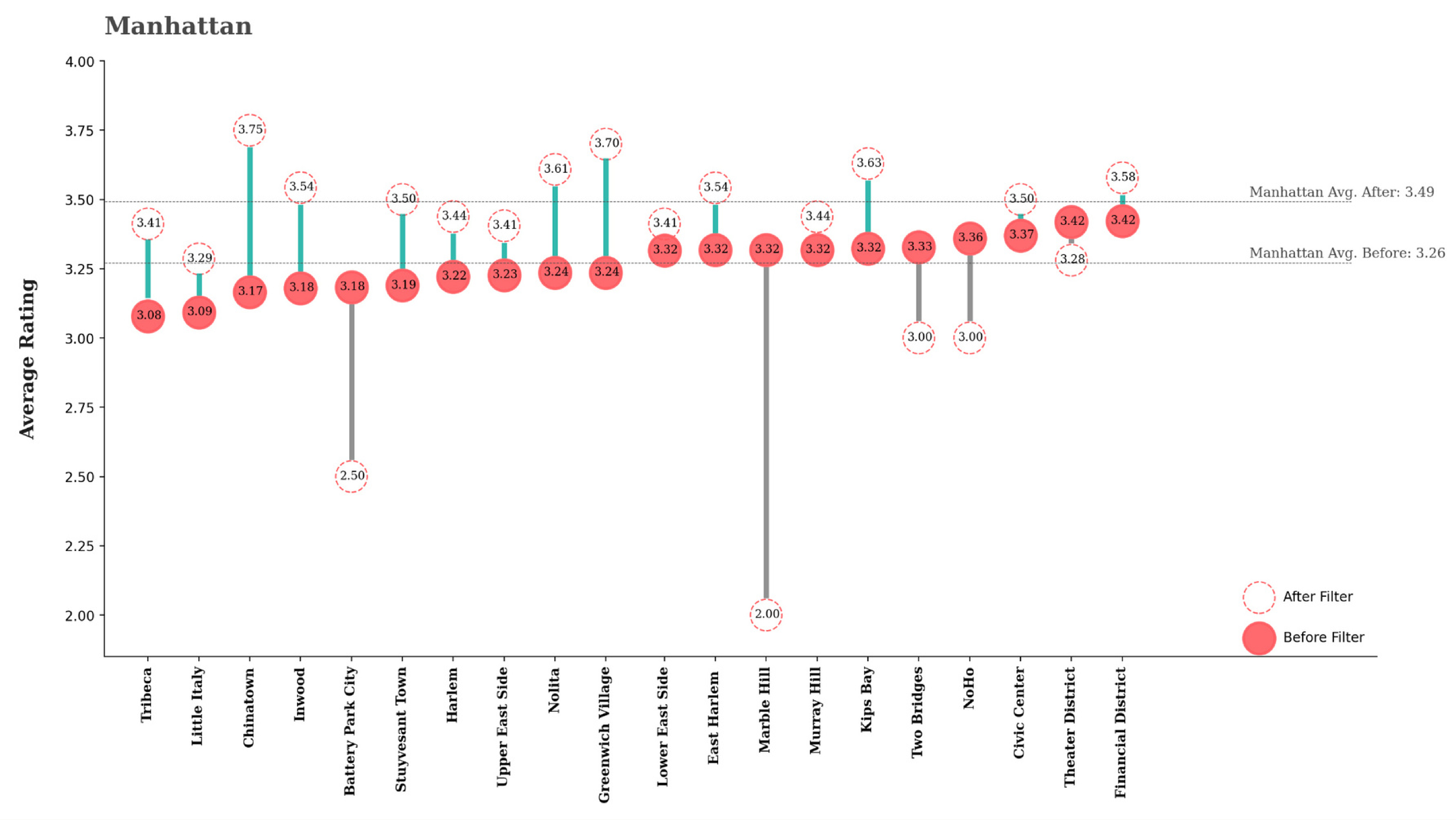
#1 Neighbourhood Showdown: *Listing Counts Across Groups*



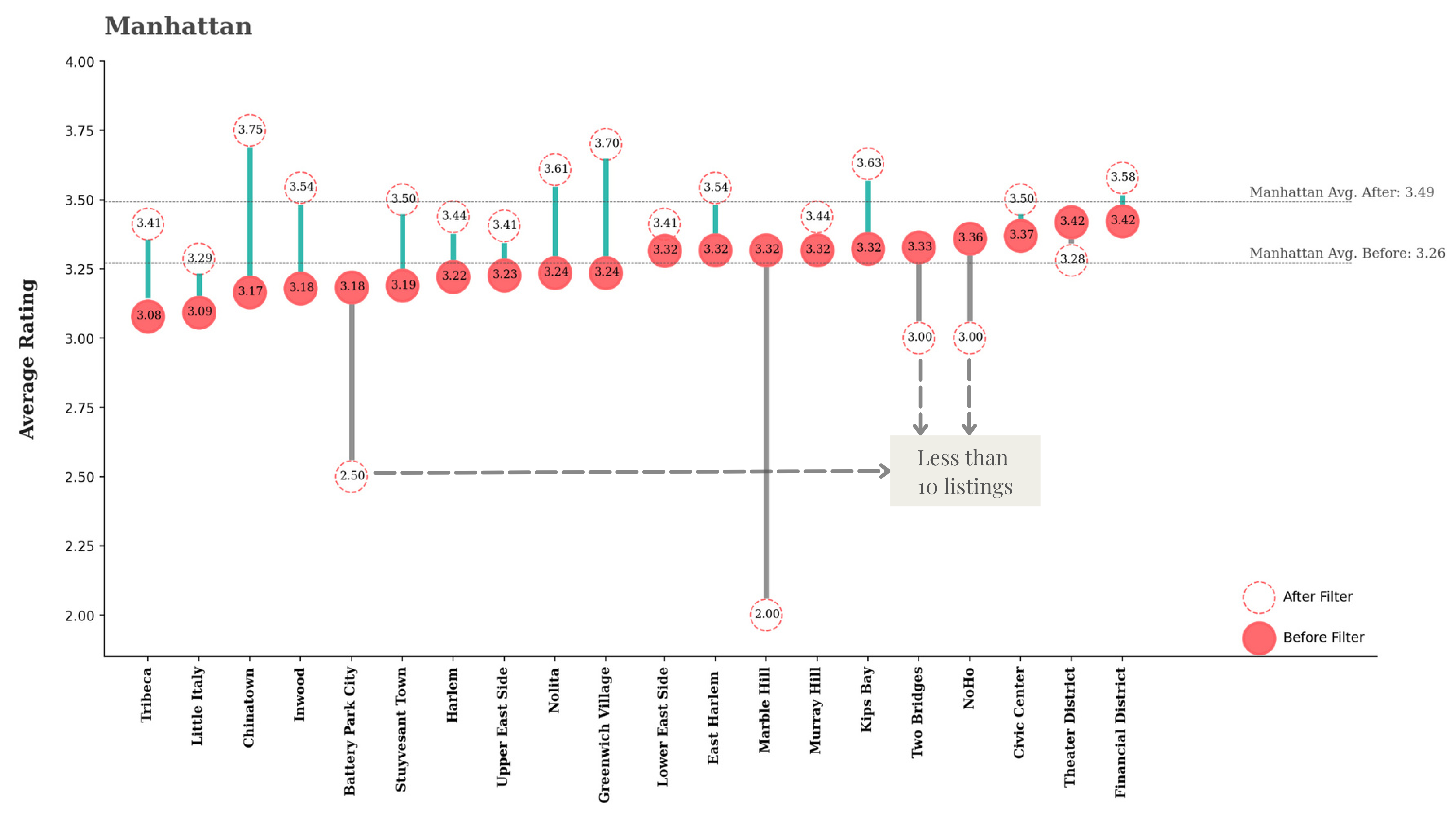
#2 Luxury Landscape: *Price Ranges by Neighbourhood Groups*



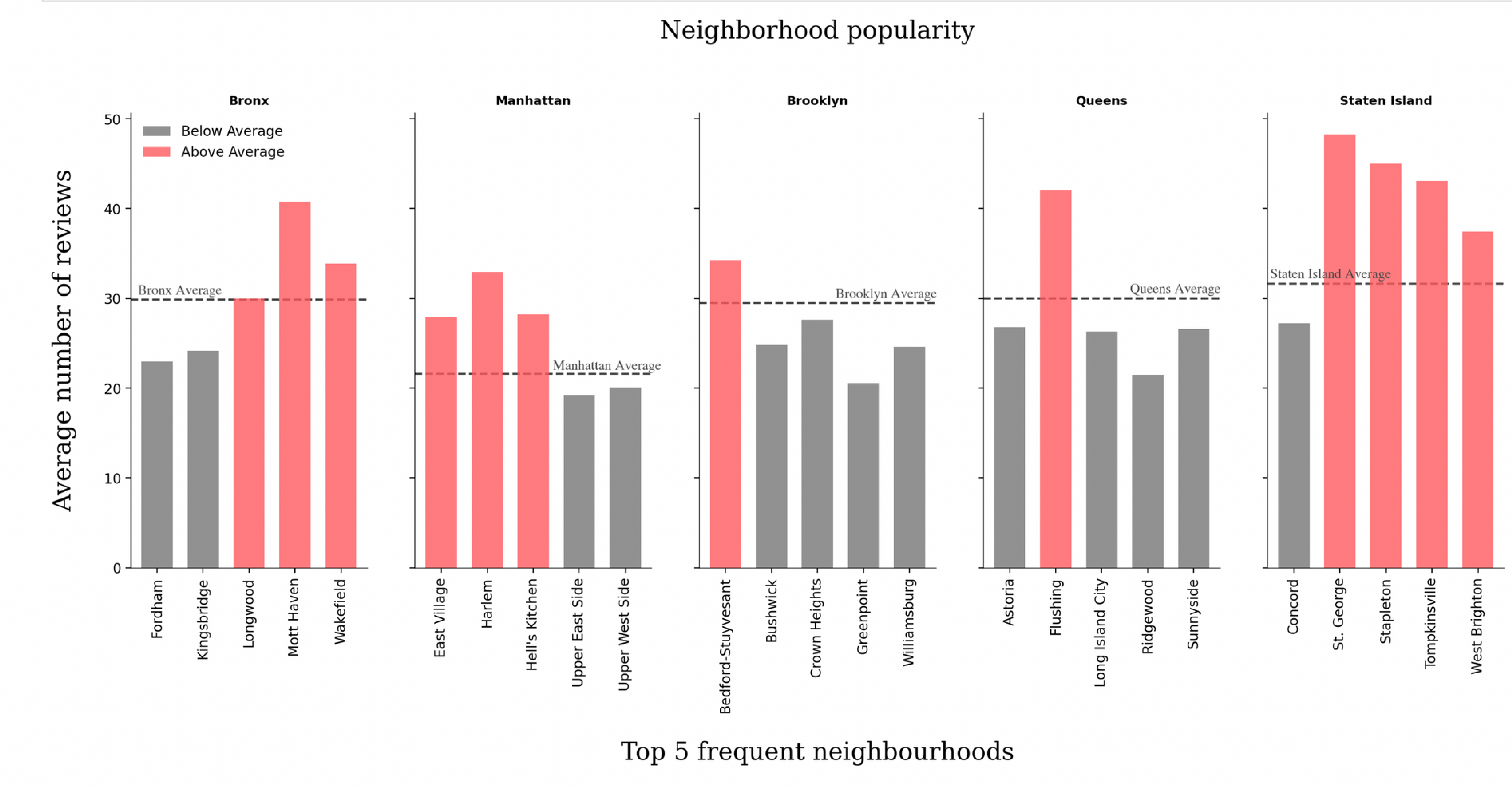
#3 Enhanced Manhattan Ratings: *Excluding Old, Long-Term, and Inactive Listings*



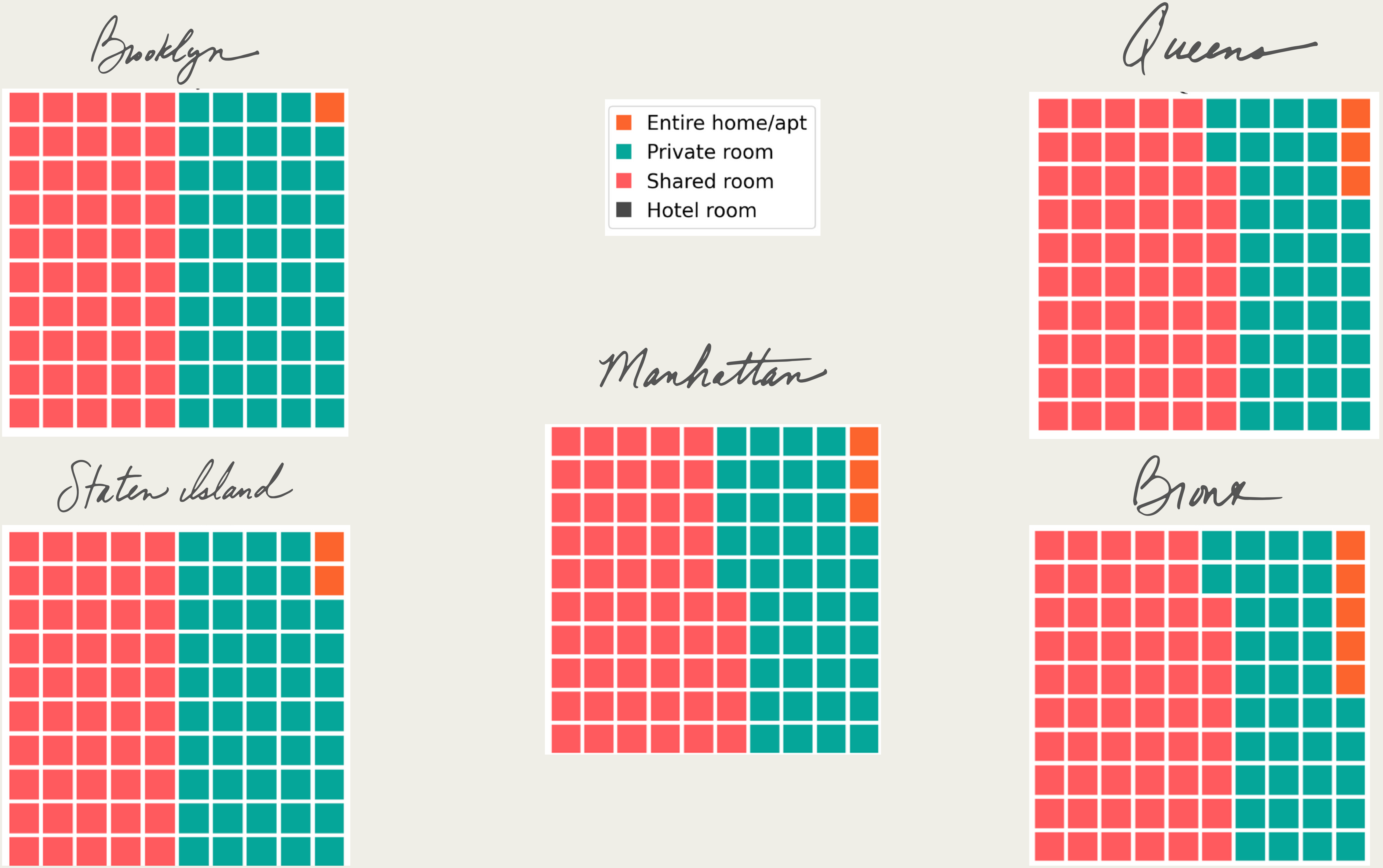
#3 Enhanced Manhattan Ratings: *Excluding Old, Long-Term, and Inactive Listings*



#4 Neighbourhood Gems: *Top Rated Hotspots Within Each Group*

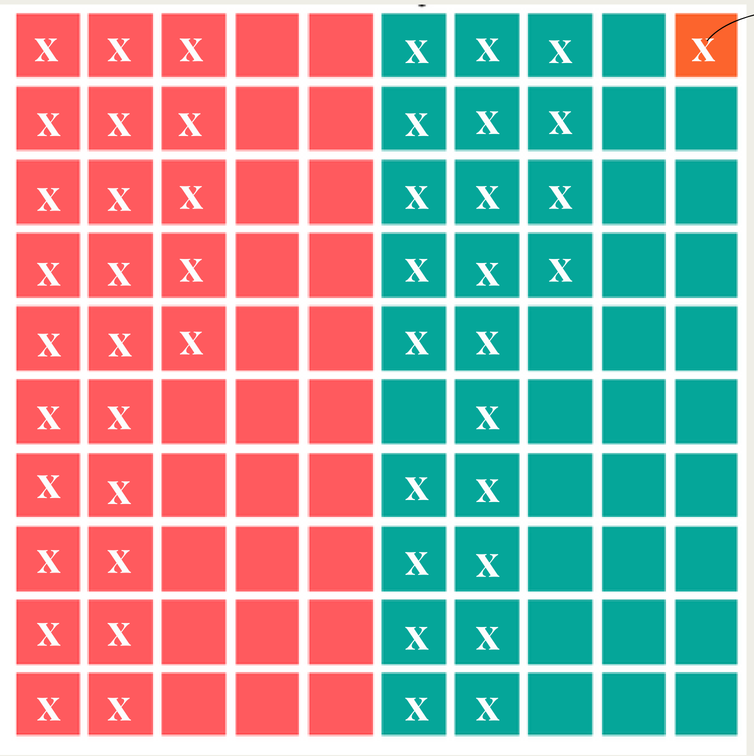


#5 Room Type: *Proportions Across Neighbourhood Groups*



#5 Room Type: *Proportions Across Neighbourhood Groups*

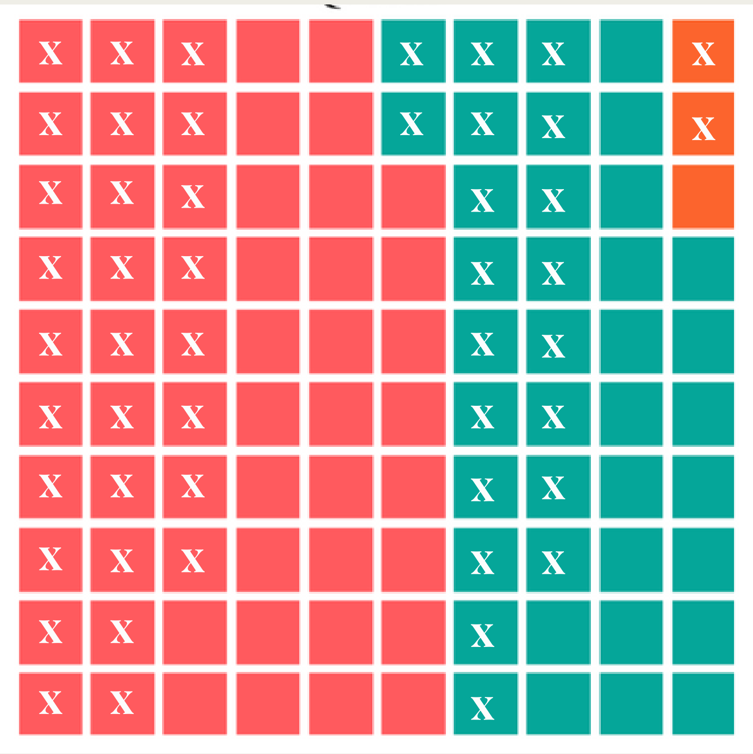
Brooklyn



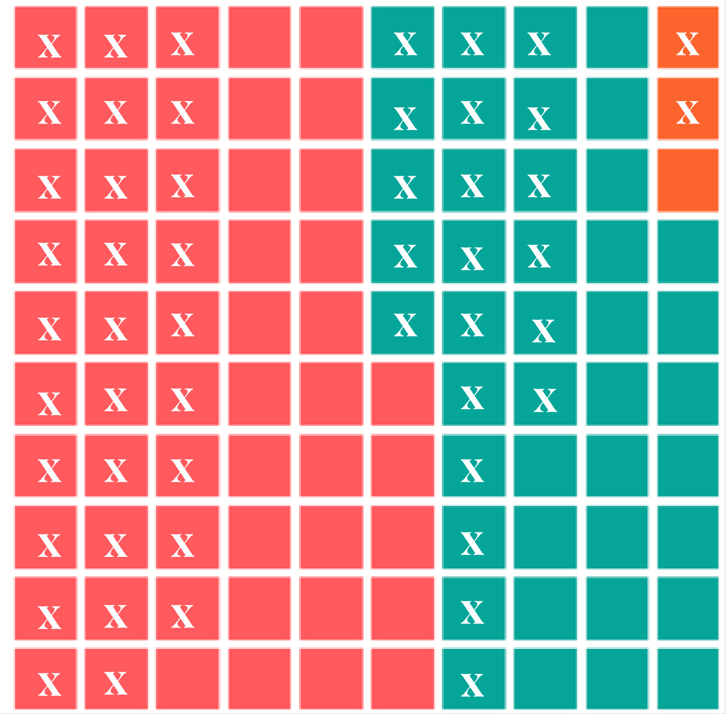
Older Than 10 Years

🟧 Entire home/apt
🟩 Private room
🟥 Shared room
⬛ Hotel room

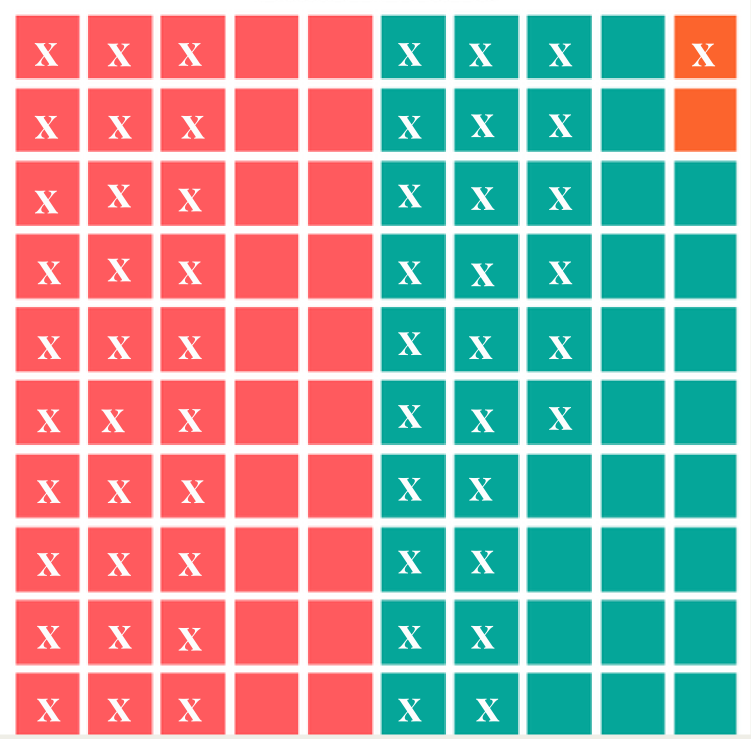
Queens



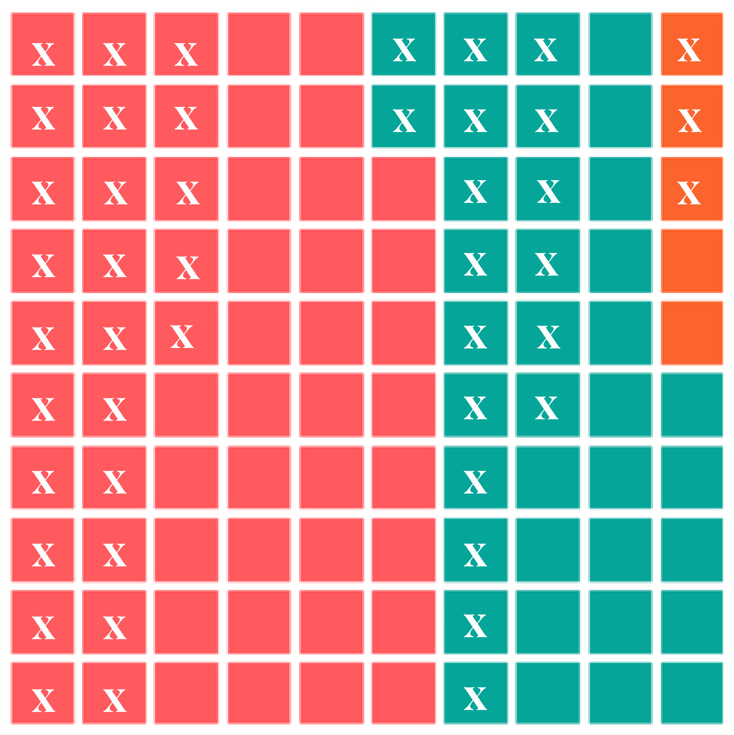
Manhattan



Staten Island

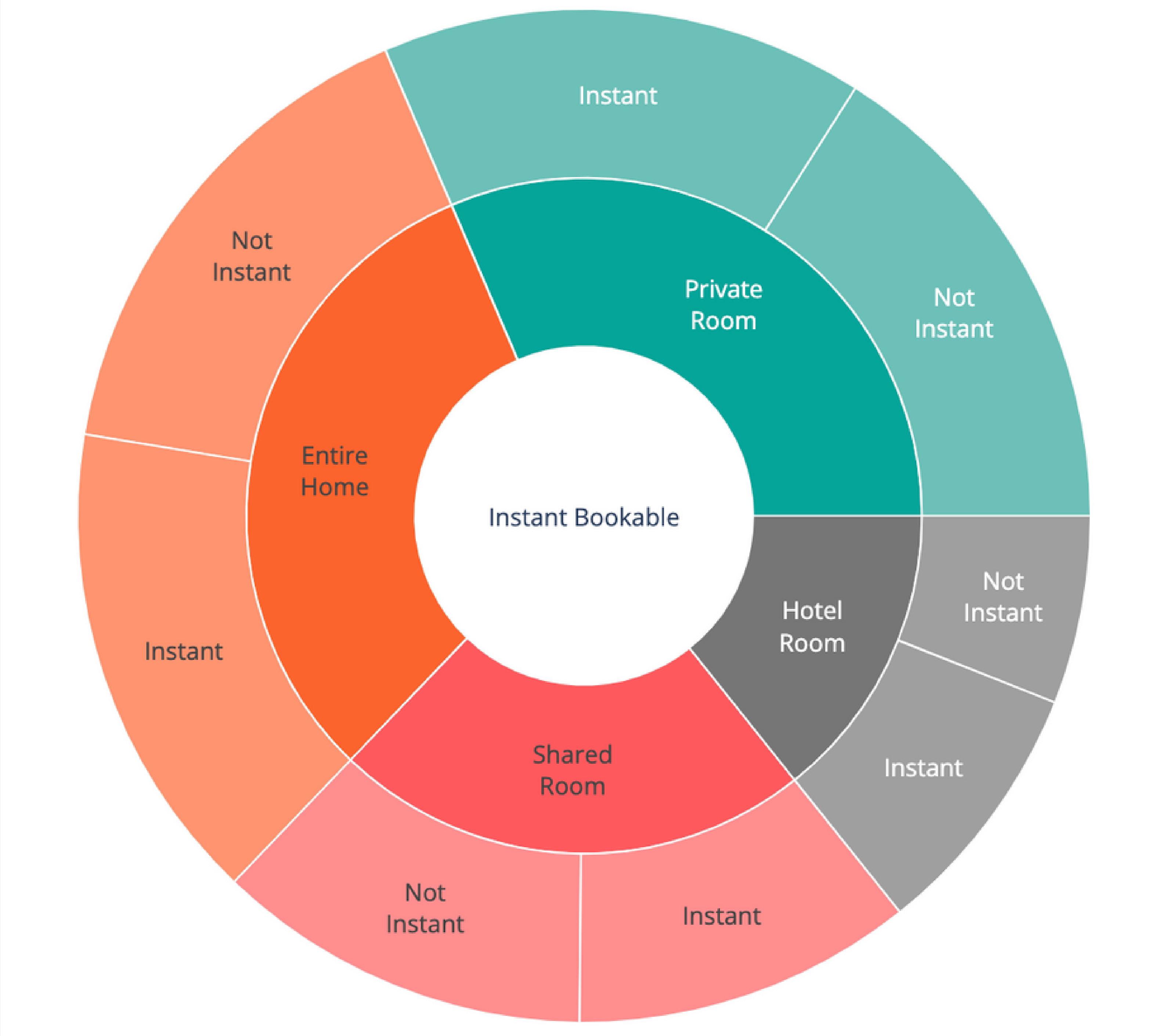


Bronx

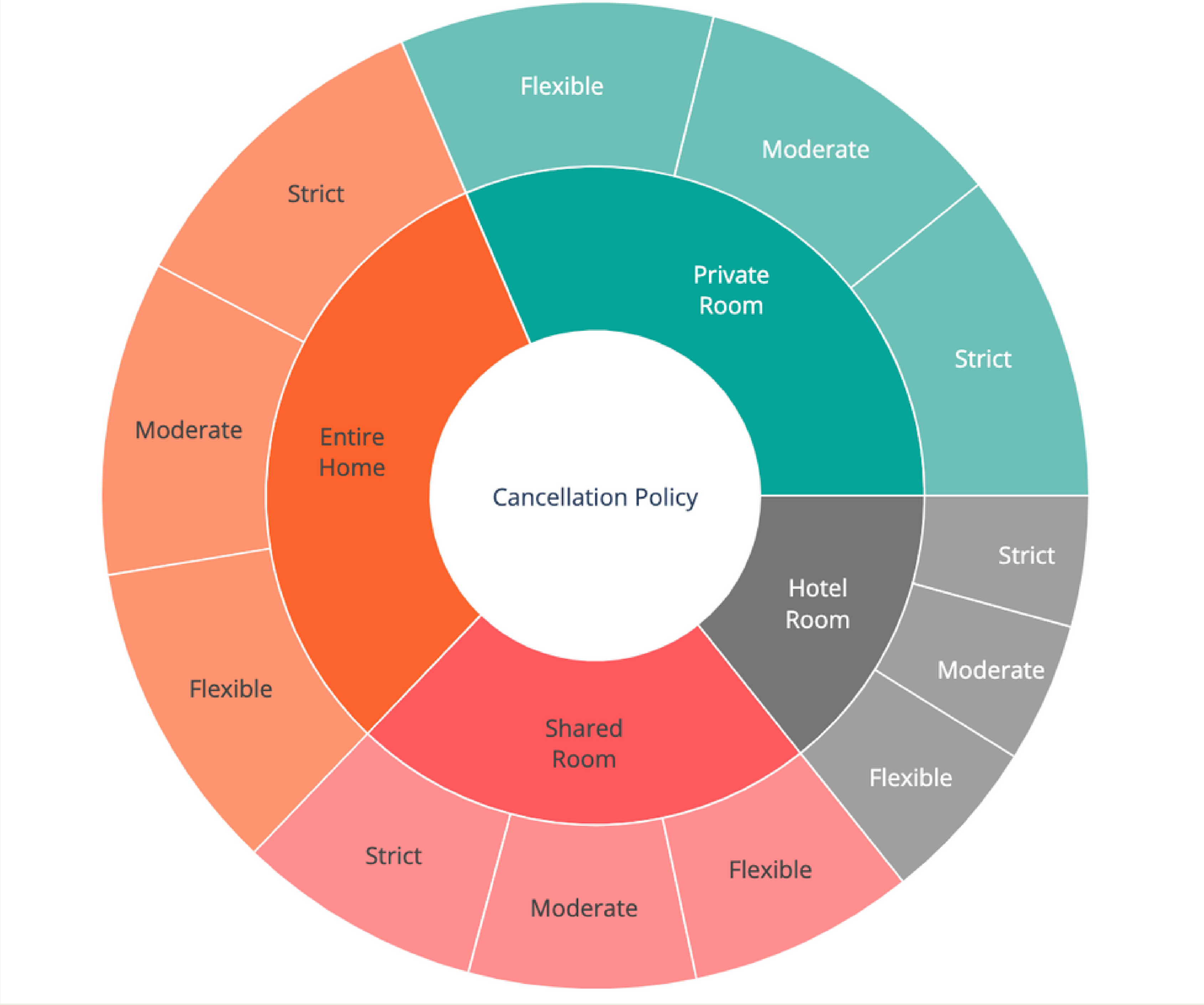


#5 Room Type: *Proportions Across Neighbourhood Groups*

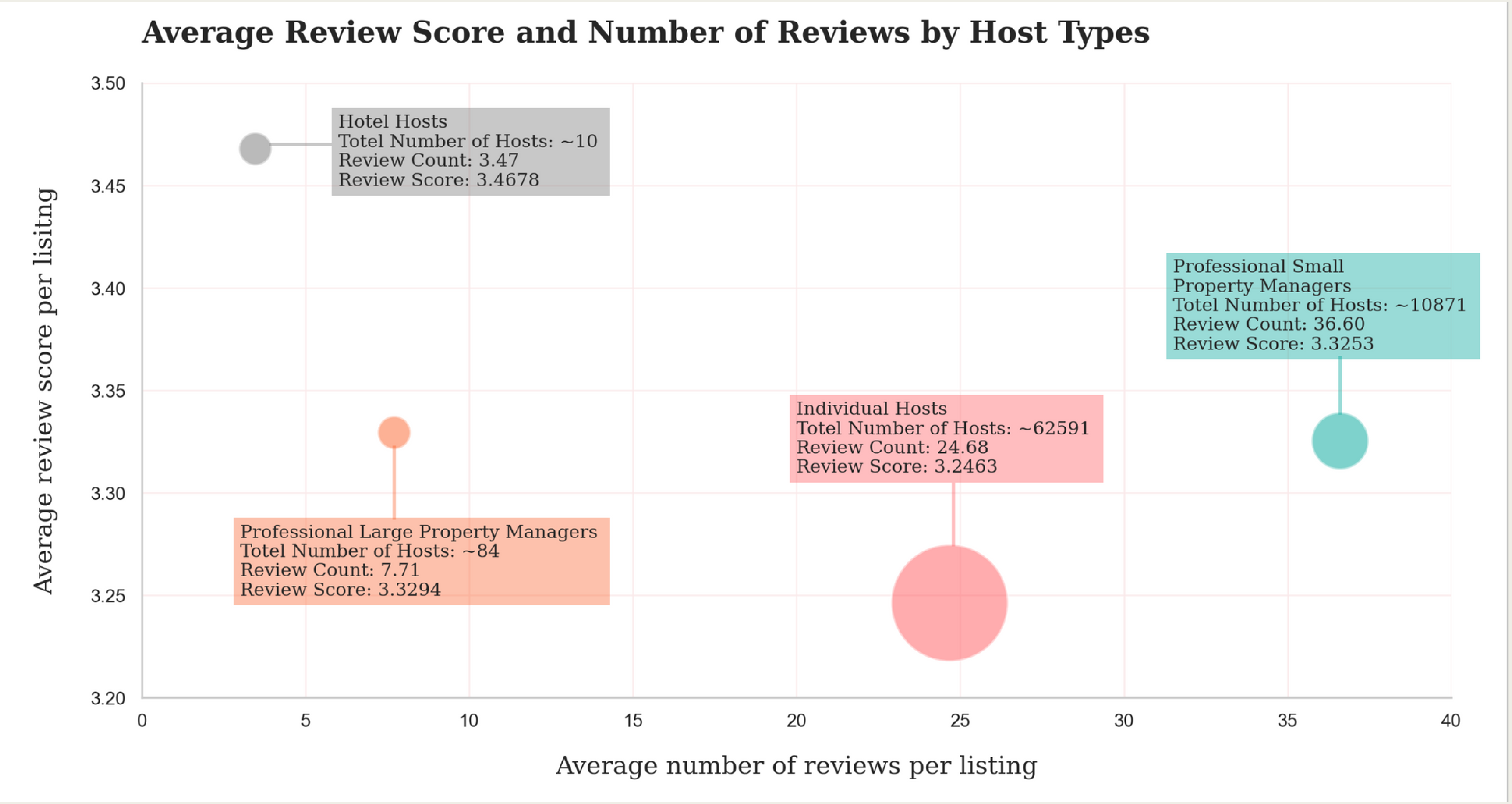
1 Instant Bookable



2 Cancellation Policy



#6 Host Analysis: *Review Scores and Customer Engagement*


Average Review Score and Number of Reviews by Host Types
Average review score per lisitng
Average number of reviews per listing
Hotel Hosts
Totel Number of Hosts: ~10
Review Count: 3.47
Review Score: 3.4678
Professional Small
Property Managers
Totel Number of Hosts: ~10871
Review Count: 36.60
Review Score: 3.3253
Individual Hosts
Totel Number of Hosts: ~62591
Review Count: 24.68
Review Score: 3.2463
Professional Large Property Managers
Totel Number of Hosts: ~84
Review Count: 7.71
Review Score: 3.3294