

# Capstone Project (EDA) Airbnb Bookings Analysis

By: Kumari Rashmi

# Agenda:

**A. Introduction**

**B. Data Preparation & Cleaning**

**C. Data Exploration and Visualization**

**D. Conclusion**

# What this presentation is about?

- Findings of EDA of Airbnb dataset for New York City
- Airbnb is an online marketplace that connects people who want to rent out their homes, with people looking for accommodation in that locale
- New York City is the most populous and expensive city in the United States and attracts a lot of travellers
- The dataset is investigated, with the goal of identifying the underlying patterns, spotting any irregularities and drawing insights

# Basic Information about the dataset

- Total 48,895 rows with 16 columns, each uniquely identified by “id” column
- Out of 16 columns, 10 are numerical & rest 6 are categorical
- Airbnb is present in all five boroughs (Brooklyn, Manhattan, Queens, Staten Island & Bronx)
- The boroughs are divided into neighbourhoods, there are 221 neighbourhoods with at least one Airbnb
- There are three types of rooms - Private , Entire home & Shared room

# Agenda

**A. Introduction**

**B. Data Preparation & Cleaning**

**C. Data Exploration and Visualization**

**D. Conclusion**

## 1. Formatting Inconsistent data type:

- date values of last review column were converted from string data type to datetime format

## 2. Handling Missing Values:

- <0.05% values are missing in "name" & "host\_name"; Since these are not relevant to our analysis, hence were ignored
- ~21% values are missing in "reviews\_per\_month" & "last\_review", following a pattern (i.e., accompanied by zero number of reviews), accordingly missing values in reviews\_per\_month column were replaced by 0

### 3. Handling Data Irregularities/Outliers:

a. \$0 Price:

- 11 Airbnb have \$0 price (of which 9 are in Brooklyn), this does not make any sense, as nobody would offer their property free of cost.
- To not lose any data, \$0 prices were substituted by median prices of respective boroughs.

b. >\$750 Price:

- Only 1% Airbnb with price >\$750. Since expensive properties are possible (i.e., basis lavishness, amenities, location etc.), so these are natural outliers.
- However, these outliers will affect the insights about price comparisons, therefore Median prices were considered instead of Mean prices.

c. Minimum nights:

- Only 1.5% Airbnb have requirement of minimum stay of >30 nights.
- As per Airbnb website & New York city rental laws, it's encouraged to host a property for 30 nights or longer. So, longer stays are natural and hence, were not treated.

d. 0-day availability:

- 36% Airbnb (17,533) are unavailable for booking during the next 365 days from data collection date.
- Airbnb allows hosts to be temporarily unavailable for booking without putting their property permanently off Airbnb. This is natural and hence, were not treated.



# Agenda

**A. Introduction**

**B. Data Preparation & Cleaning**

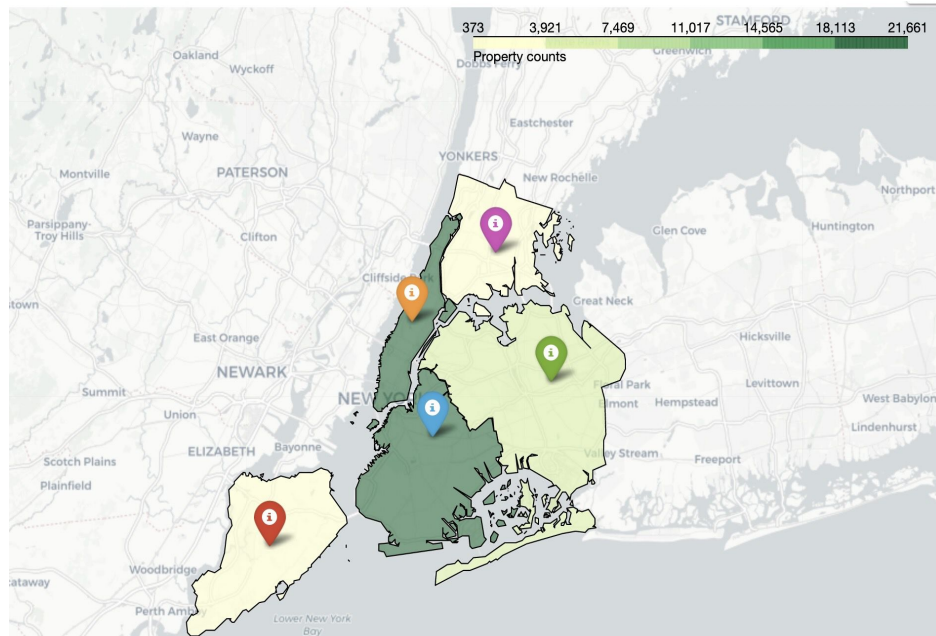
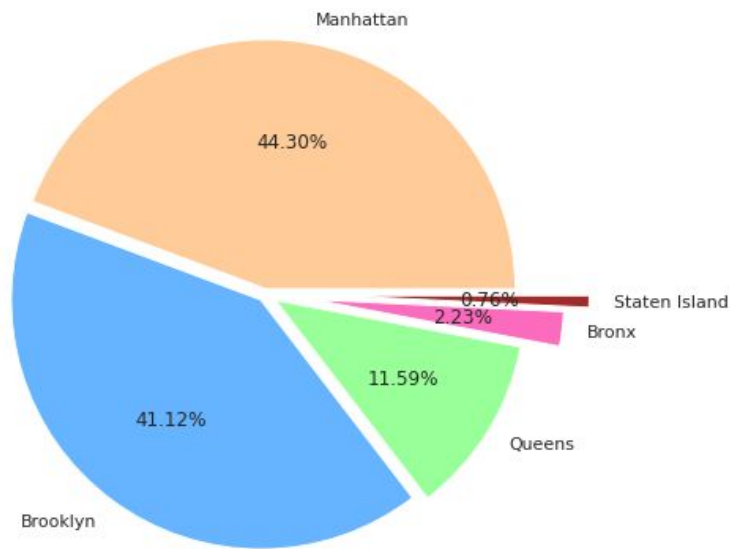
**C. Data Exploration and Visualization**

**D. Conclusion**

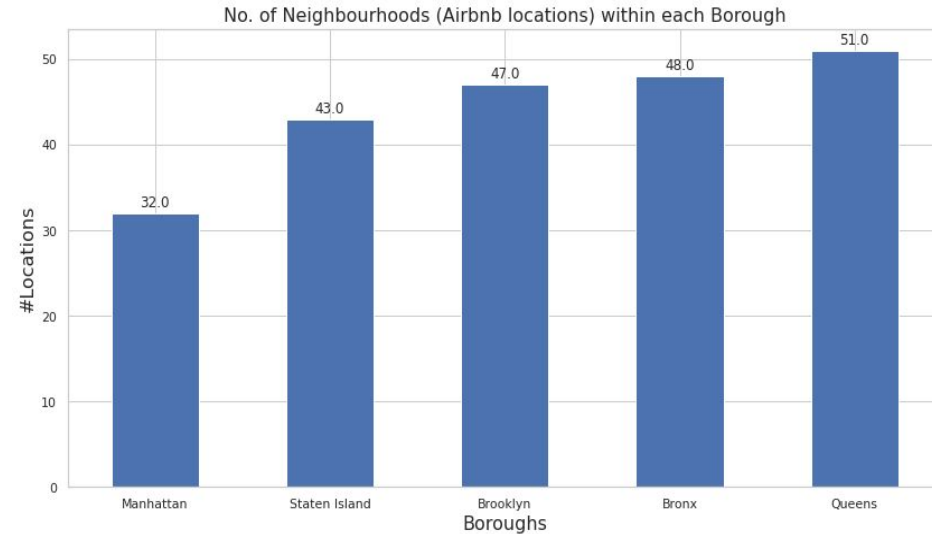
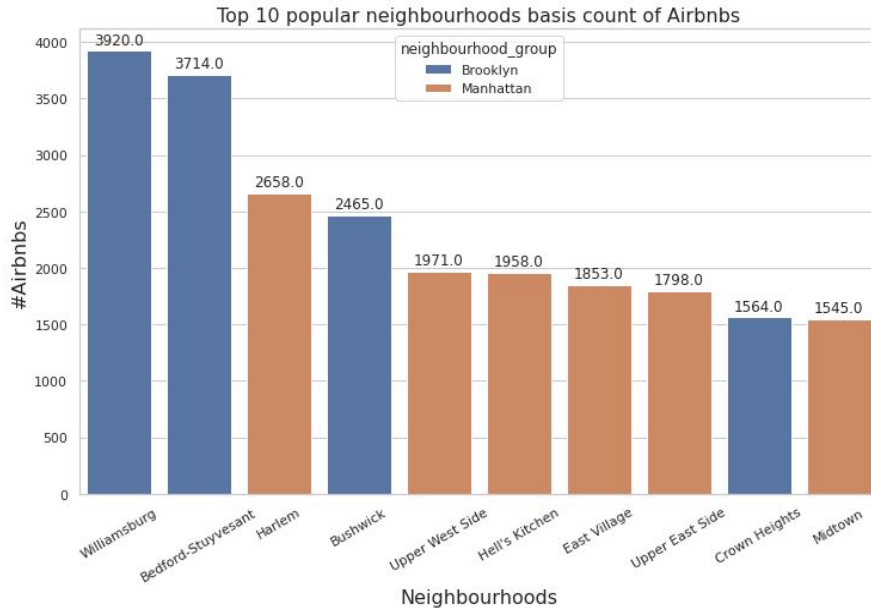
# 1. Locations: boroughs & neighborhoods

- Manhattan (21,661) has highest share of Airbnb, followed by Brooklyn (20,104)
- Staten Island has the lowest share

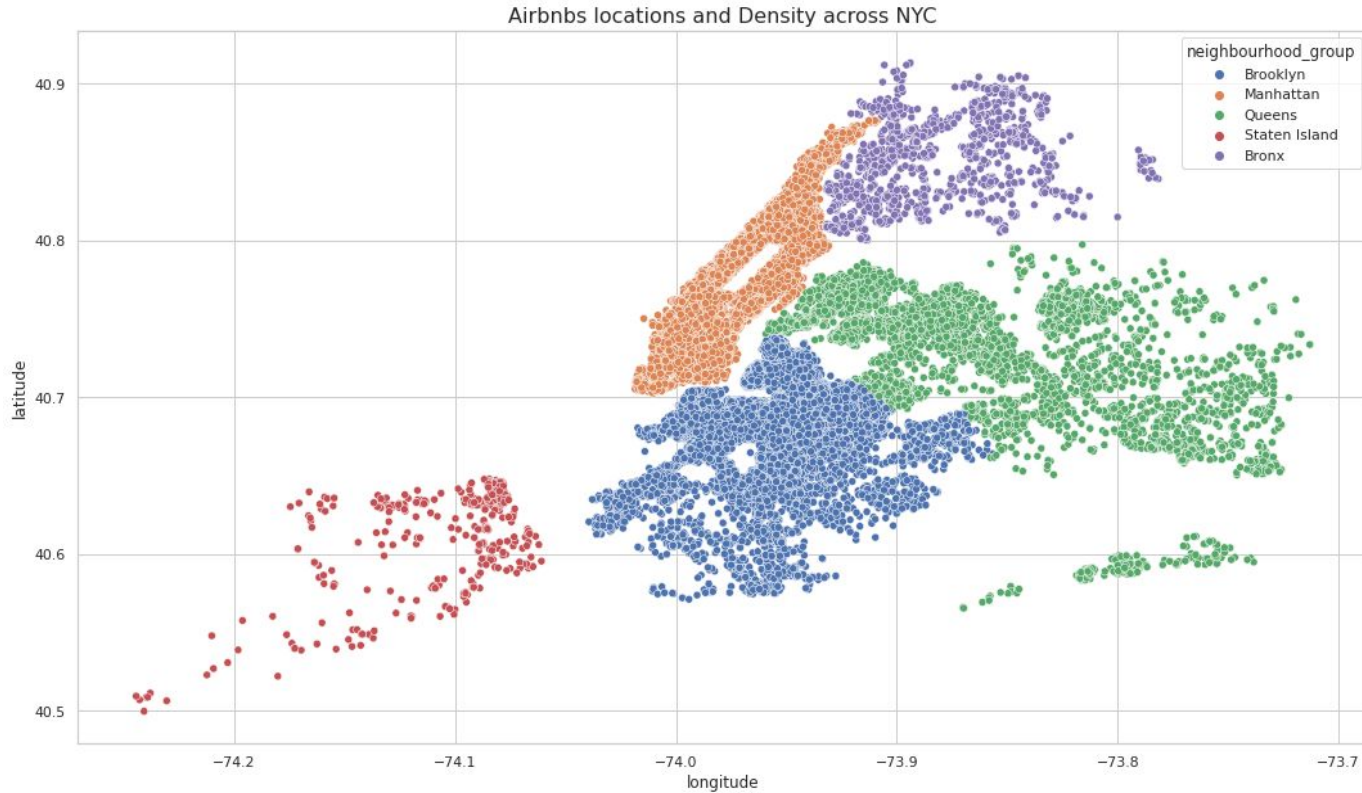
Distribution of Airbnbs across 5 NYC Boroughs



- Williamsburg, Brooklyn is the neighborhood with the highest number of Airbnb
- Manhattan and Brooklyn contain all the top 10 neighborhoods by count
- #Neighborhoods with Airbnb is highest in Queens & lowest in Manhattan



- Manhattan has the highest number of Airbnbs located in the fewest number of neighbourhoods. The high density of Airbnbs is evident in the below scatterplot

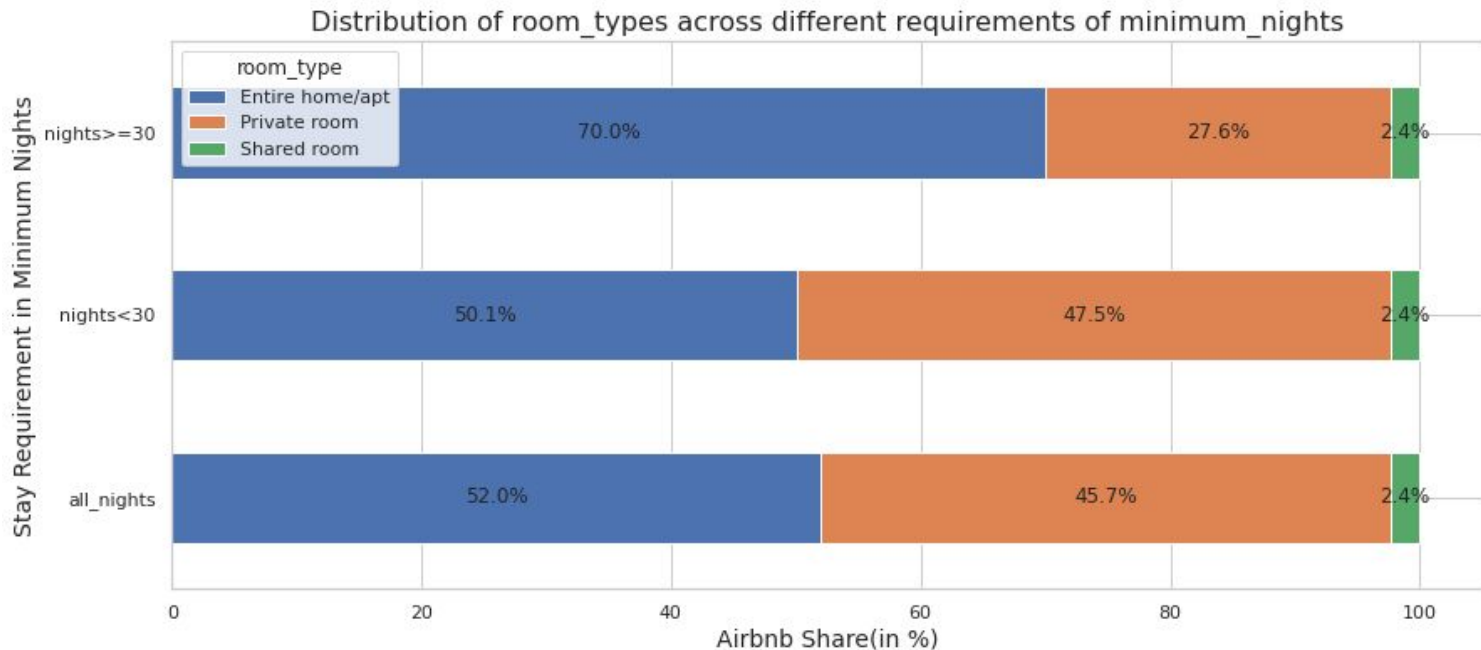


## 2. Airbnb's Minimum Stay Requirement

- Majority of Airbnb (75%) require minimum stay  $\leq 5$  nights. Short term stays are more popular
- Only 1.5% Airbnb require minimum stay  $>30$  nights. Airbnb has no restriction on upper limit of booking days in NYC. It is also encouraged to host for 30 nights or longer as per Airbnb website & NYC rental laws

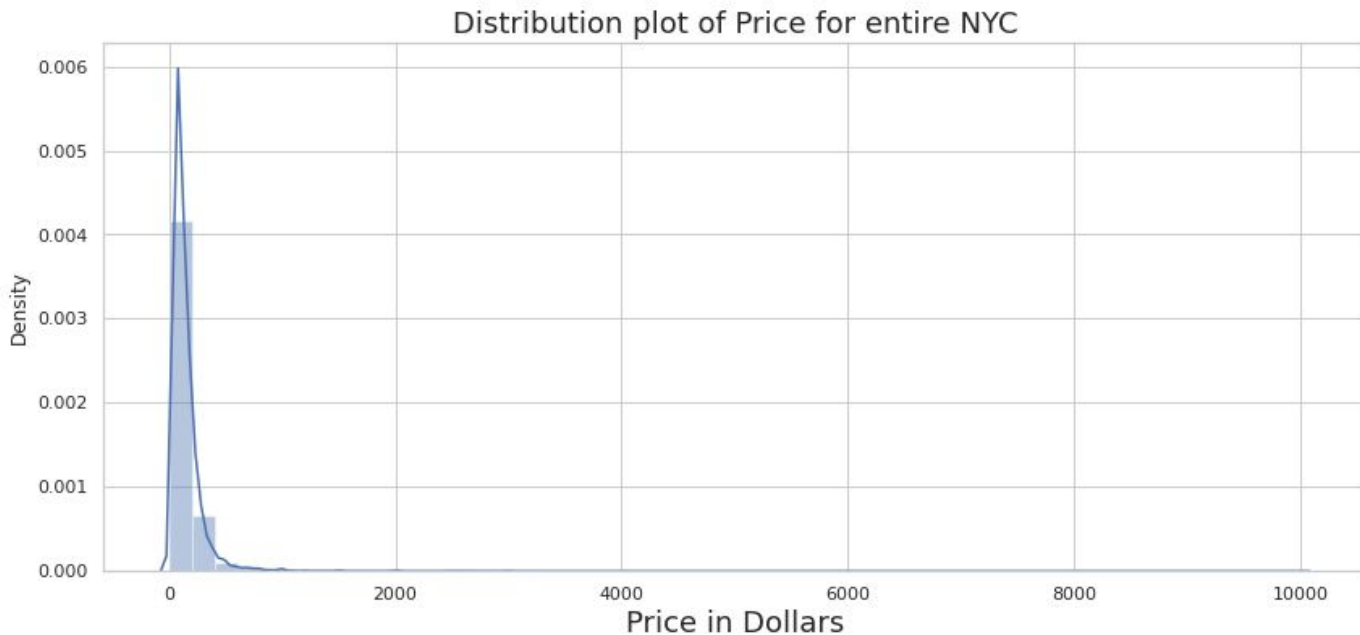
S. No.	Required Minimum nights for Booking	#Airbnb (%)
1	nights $\leq 10$	86.4 %
2	nights $\leq 30$	98.5 %
3	nights $> 30$	1.5 %

- As per NYC rental law, one can only rent out a portion of apartment for  $<30$  days, only if the host lives there
- Also as per dataset, "Entire home/apt" type Airbnb is significantly more (70%) where stay requirement is  $\geq 30$  nights. This is in accordance with NYC rental law.



### 3. Untangling the Price Trends

- Due to outliers, the distribution plot of prices is highly skewed with long tail on the right (positively skewed)
- Mean and Median prices of an Airbnb in entire NYC are \$153 and \$106 respectively



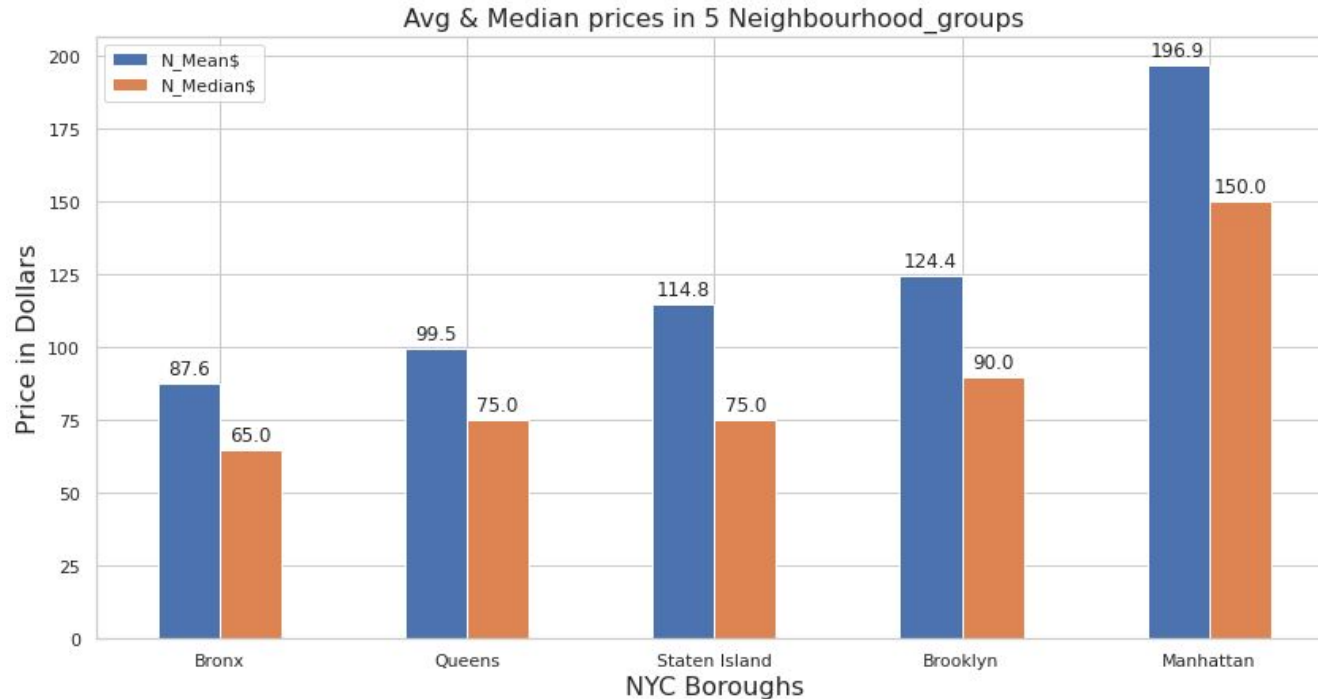


- Above boxplots indicate that majority of expensive Airbnbs are either in Manhattan or in Brooklyn

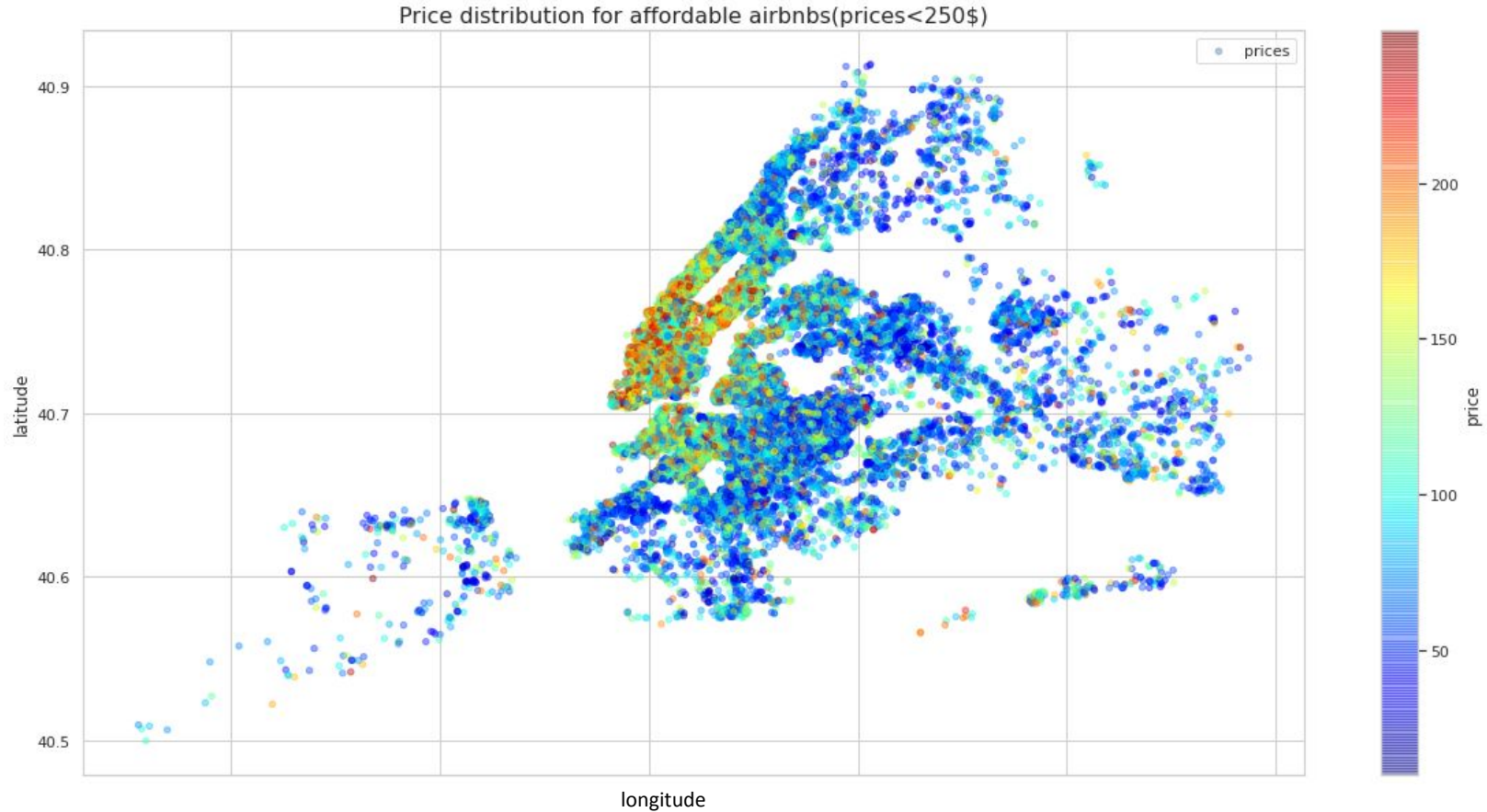


## Most Expensive Vs Affordable borough

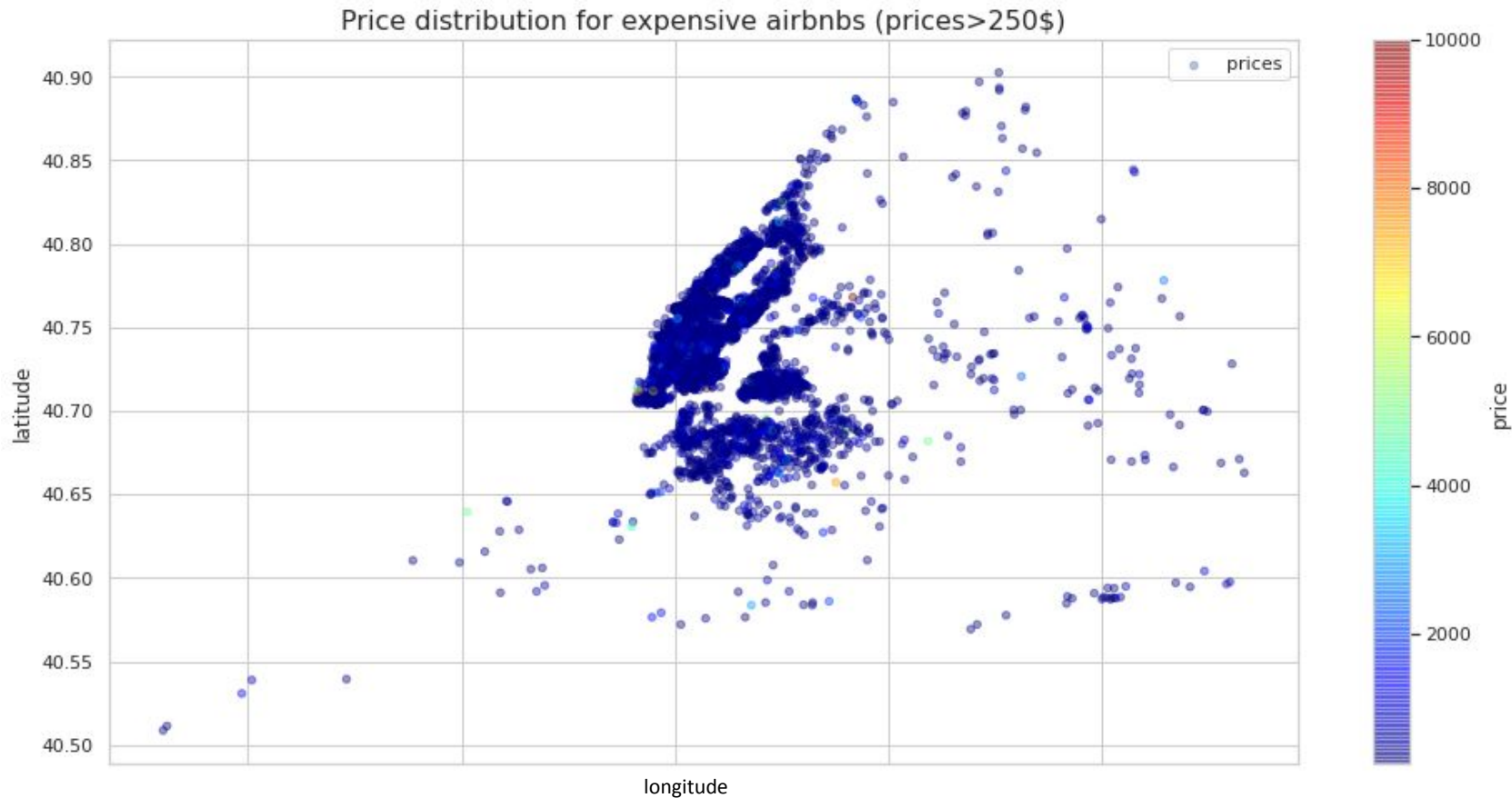
- Manhattan is the most expensive borough, followed by Brooklyn
- Bronx borough is the most affordable one



# Affordable Airbnbs : Scatter Plot



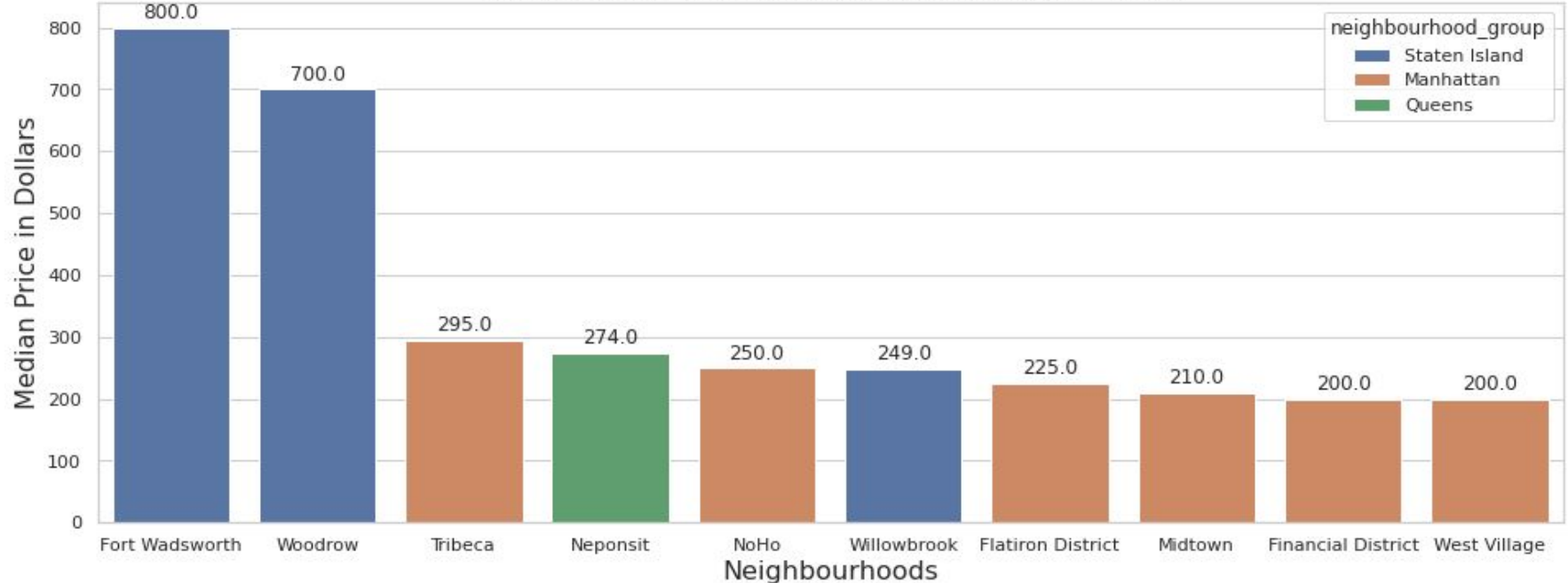
# Expensive Airbnbs : Scatter Plot



## Top 10 Expensive Neighborhoods

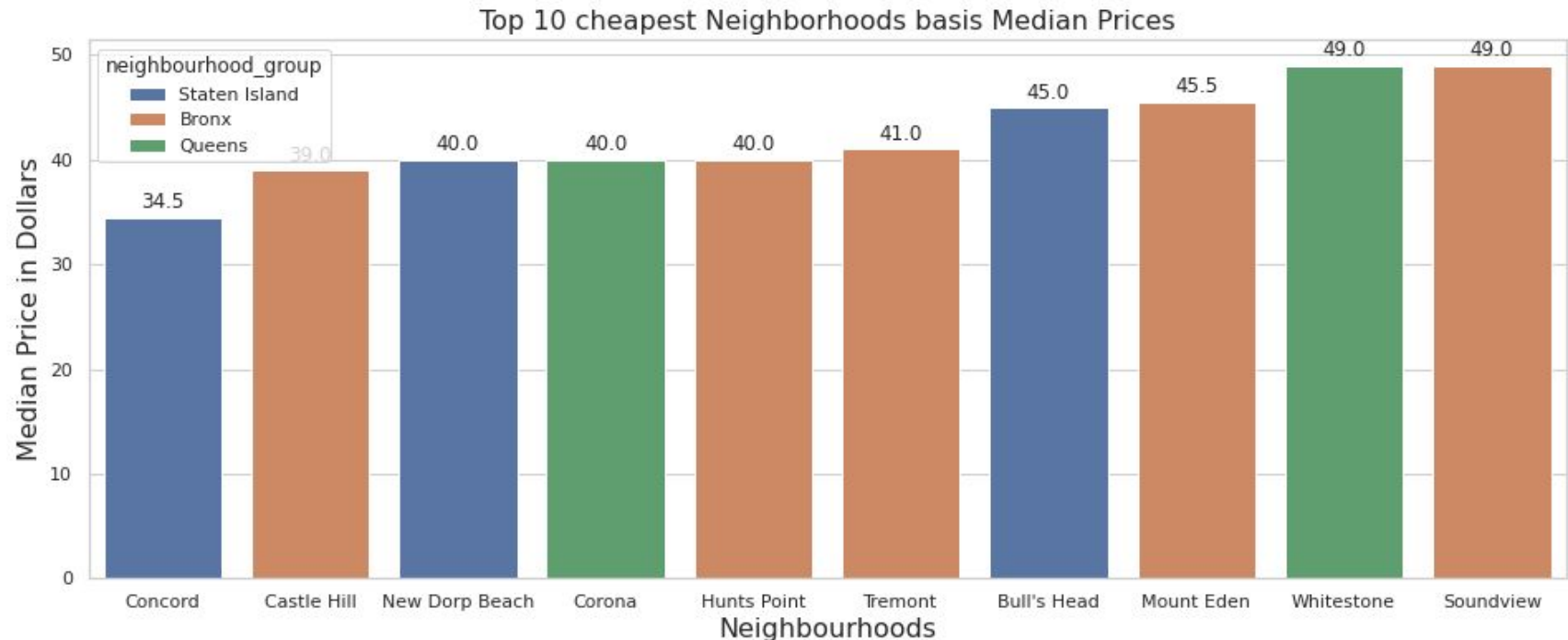
- Fort Wadsworth of Staten island is the most expensive Neighbourhood
- 6 out of the top 10 expensive neighbourhoods are in Manhattan borough

Top 10 expensive Neighborhoods basis Median Prices



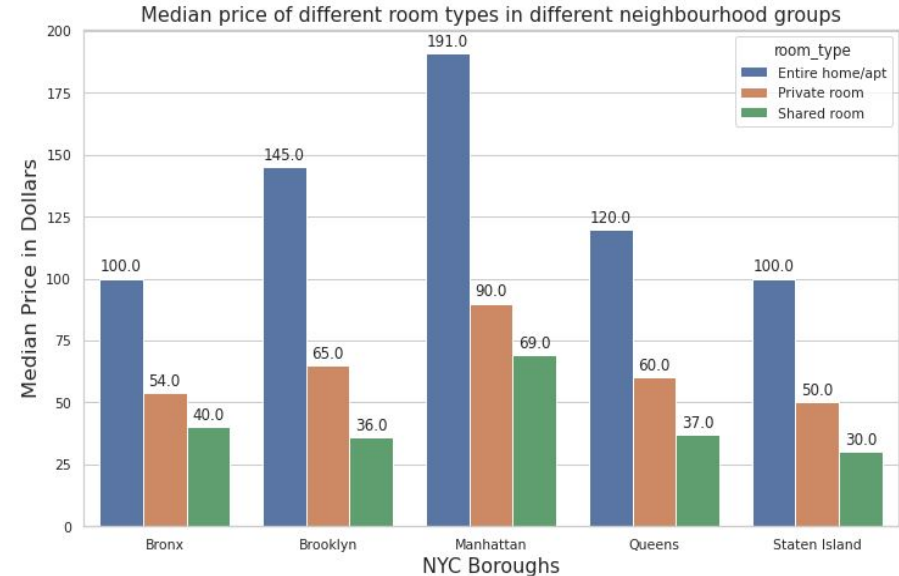
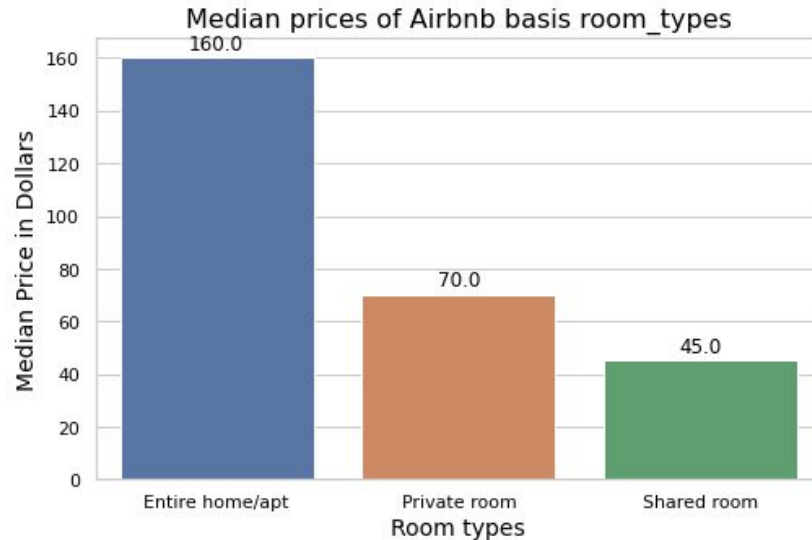
## Top 10 Most Affordable Neighborhoods

- Concord of Staten Island is the most affordable neighbourhood
- 5 out of the top 10 most affordable neighbourhoods are in Bronx borough



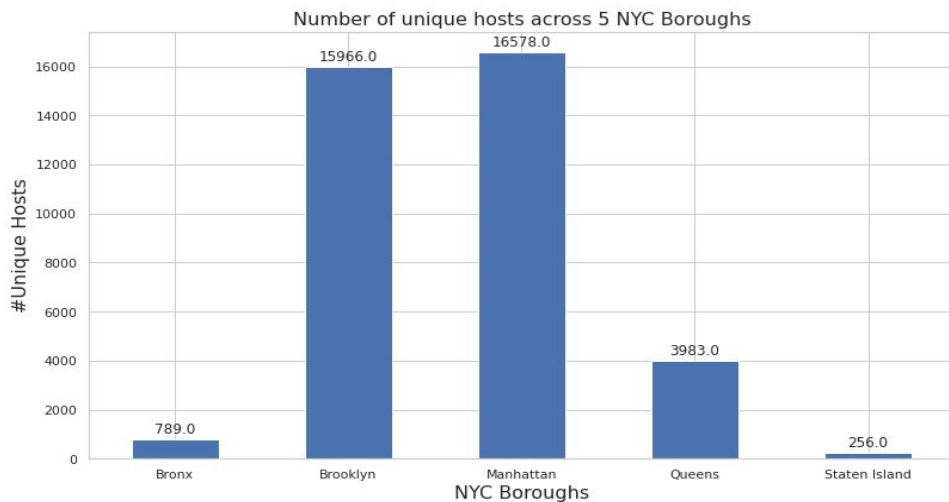
## Expensive Vs Affordable type of Room

- Entire home type is the most expensive Airbnb, while "shared room" type is the most affordable across NYC, and also within each borough
- The most expensive Airbnb is of Entire home/apt type in Manhattan, and the most affordable is of Shared room type in Staten Island



## 4. Understanding the hosts

- Total 37,457 unique hosts; #unique hosts is highest in Manhattan & Brooklyn
- 5154 hosts have multiple Airbnb, of which 642 hosts have Airbnb at more than one location.
- Out of above 642 hosts, 110 Hosts have Airbnb in more than one borough & all are of "Private" room type.





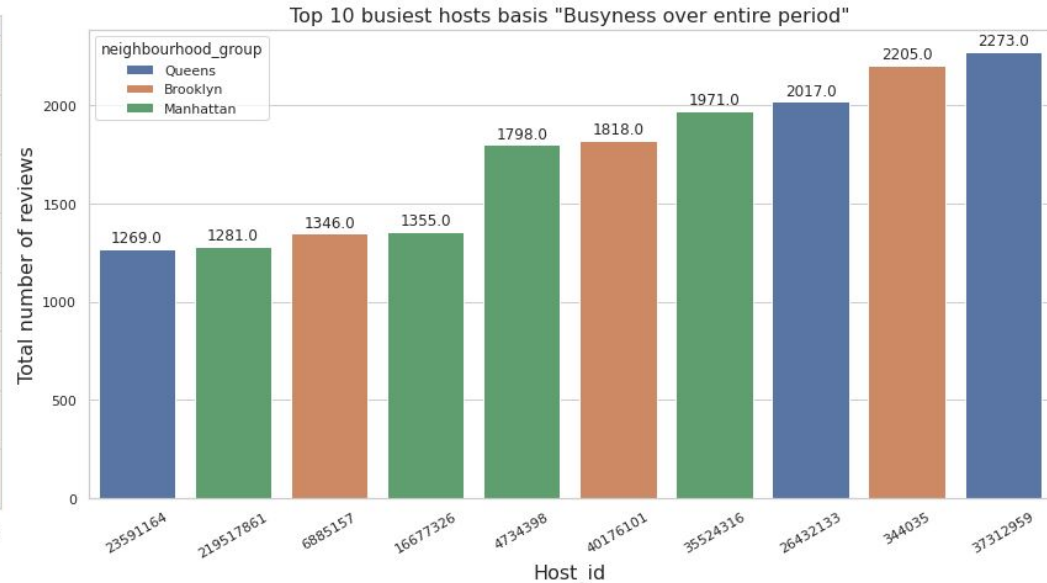
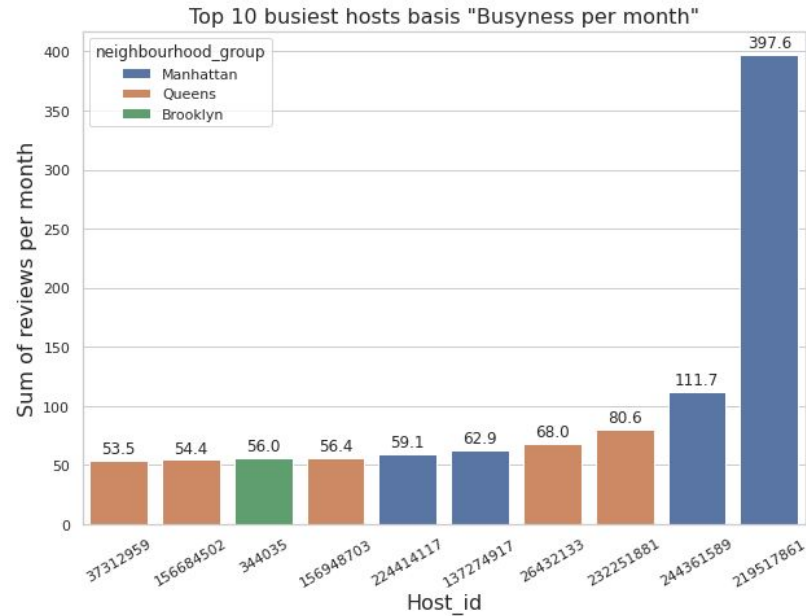
Assumption: majority of guests give reviews

- Total of reviews\_per\_month was taken as an indicator of “busyness per month”. Total number\_of\_reviews was taken as an indicator of “busyness over entire period”
- Basis “busyness per month”, Host {219517861, Sonder(NYC)} ranks highest. They also have the highest number of property listings (327) in Manhattan, mainly in Financial District, Murray Hill, Theater districts neighbourhoods, which attracts a lot of traffic
- Basis “busyness over entire period”, Host{37312959, Maya} ranks the highest. They have 5 Airbnb near LaGuardia airport in Queens, of Private room type and with minimum stay requirement of 1 night. These all are guest friendly points leading to more traffic and so more reviews



## Top 10 Busiest Hosts

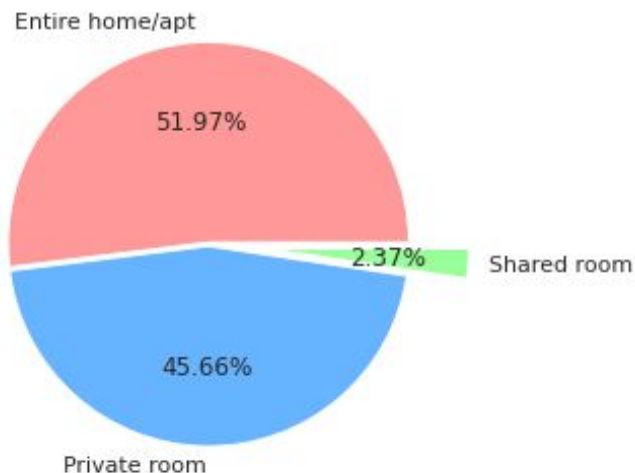
- Manhattan & Queens has most of the busiest hosts basis “Busyness per month”
- Queens, Brooklyn & Manhattan has all the top 10 busiest hosts basis “Busyness over entire period”



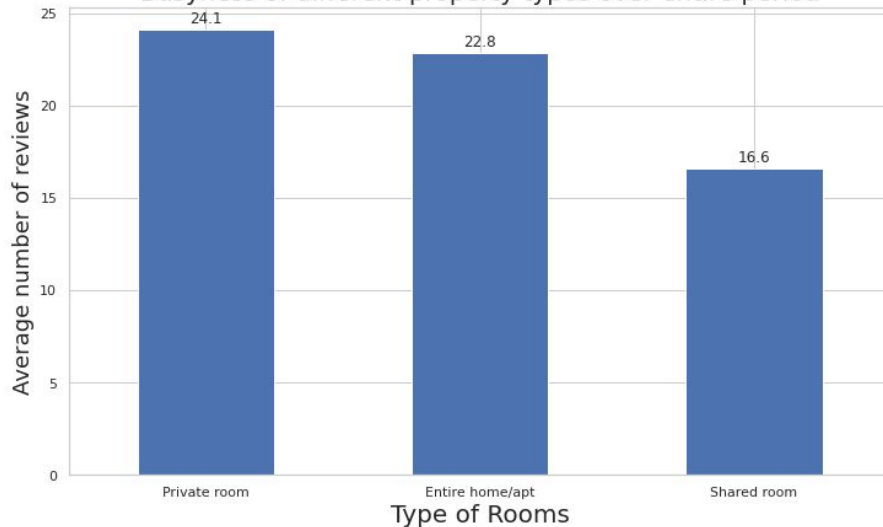
## 5. Property types

- Majority of Airbnb are either of "Entire home/apt" or "Private room" type; Airbnb of "Shared room" type are the least
- Basis average number\_of\_reviews, "Private room" type Airbnb is the busiest, followed by "Entire Home/apt." and "shared room"

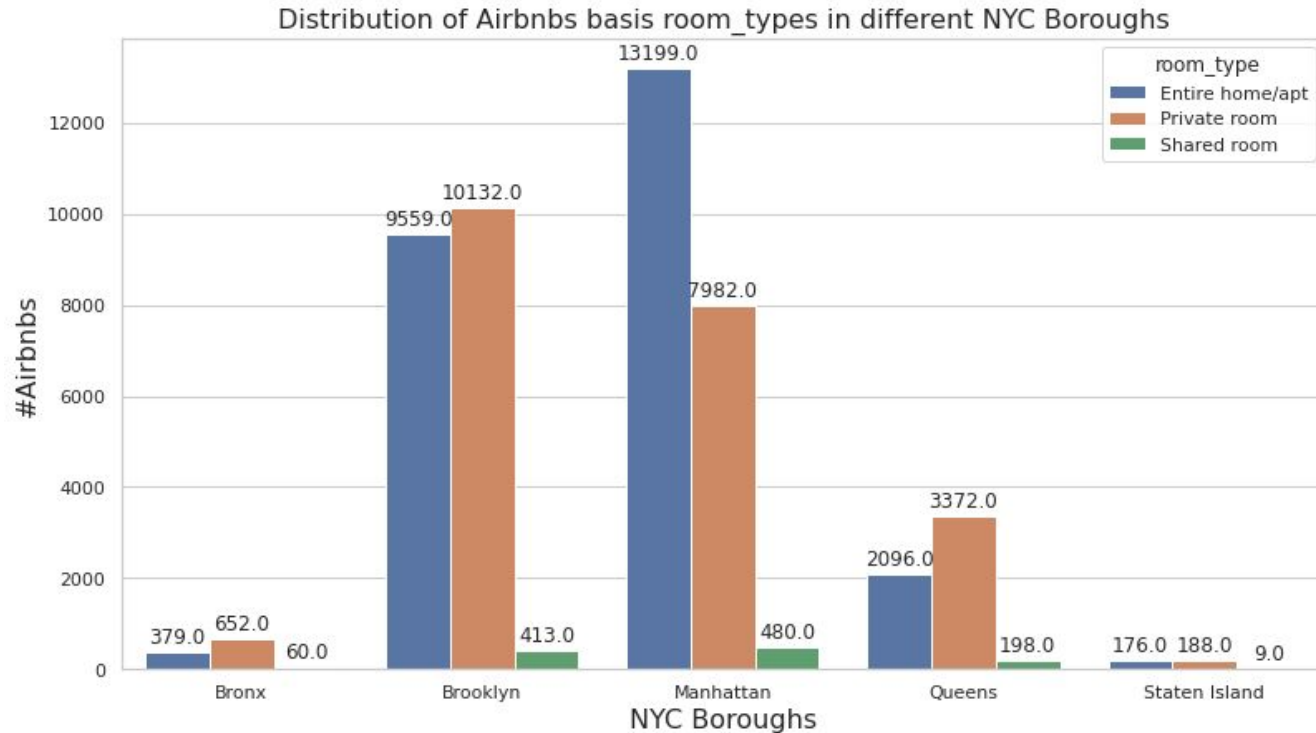
Distribution of Airbnb basis room types



Busyness of different property types over entire period



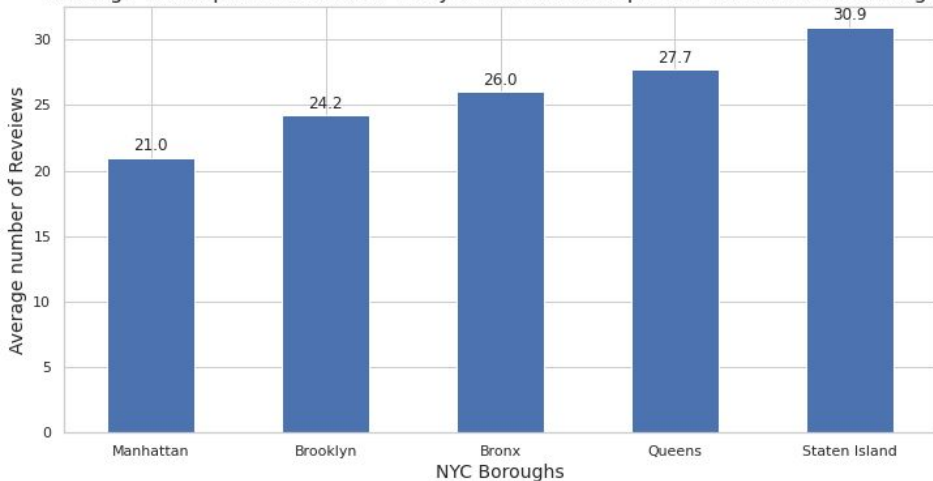
- In all 5 NYC boroughs except Manhattan, majority of Airbnb are of Private room type
- In Manhattan, Entire home/apt type Airbnb are significantly more than Private room type



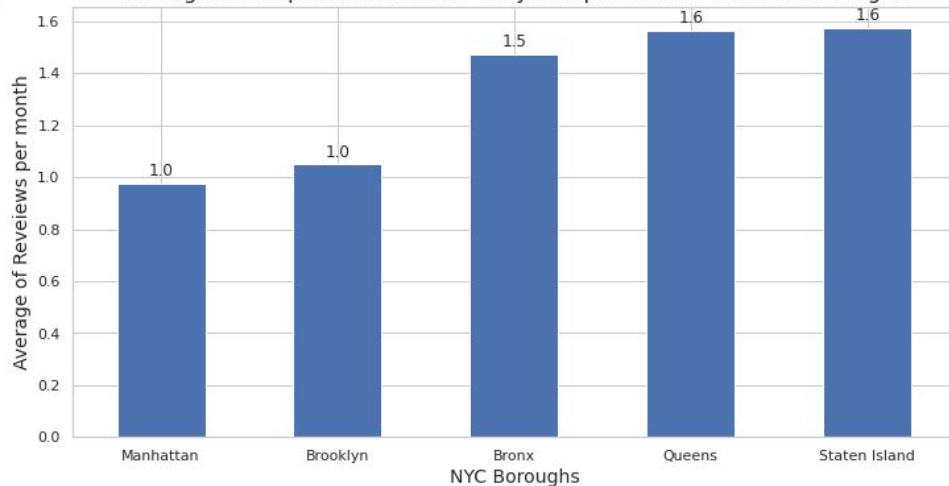
## 6. Busyness of an Airbnb

- Basis “Busyness per month” and “Busyness over entire period”, an Airbnb located in Staten Island is the busiest, with the highest average traffic per Airbnb
- This might be due to difference in market of the boroughs, high margin vs high volume business models

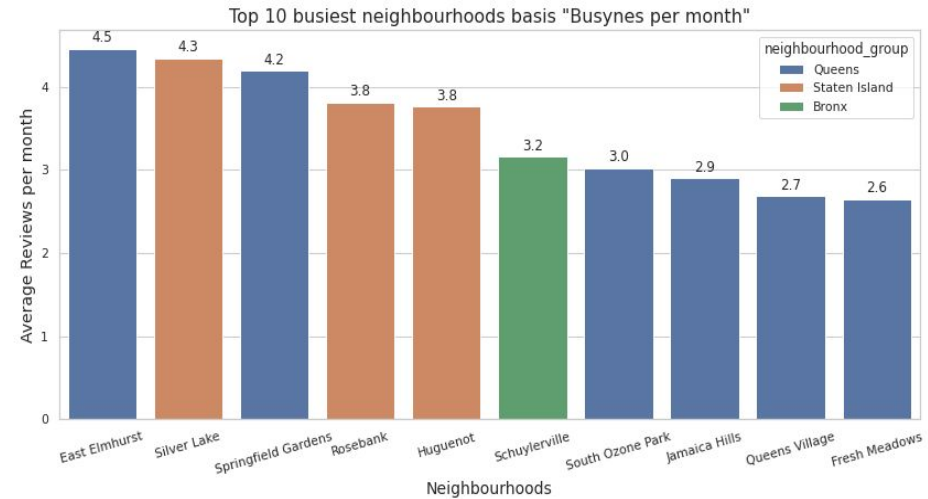
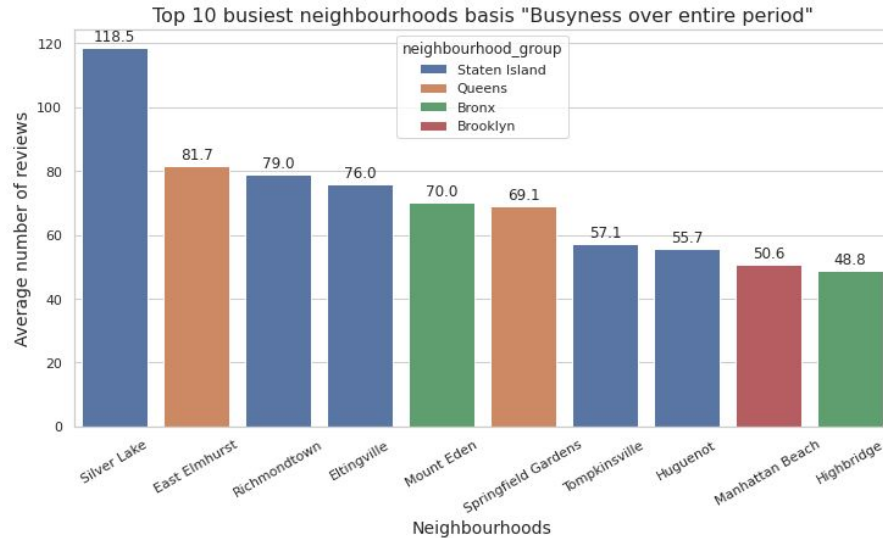
Average Traffic per Airbnb basis “Busyness over entire period” across 5 NYC Boroughs



Average Traffic per Airbnb basis “Busyness per month” across 5 Boroughs

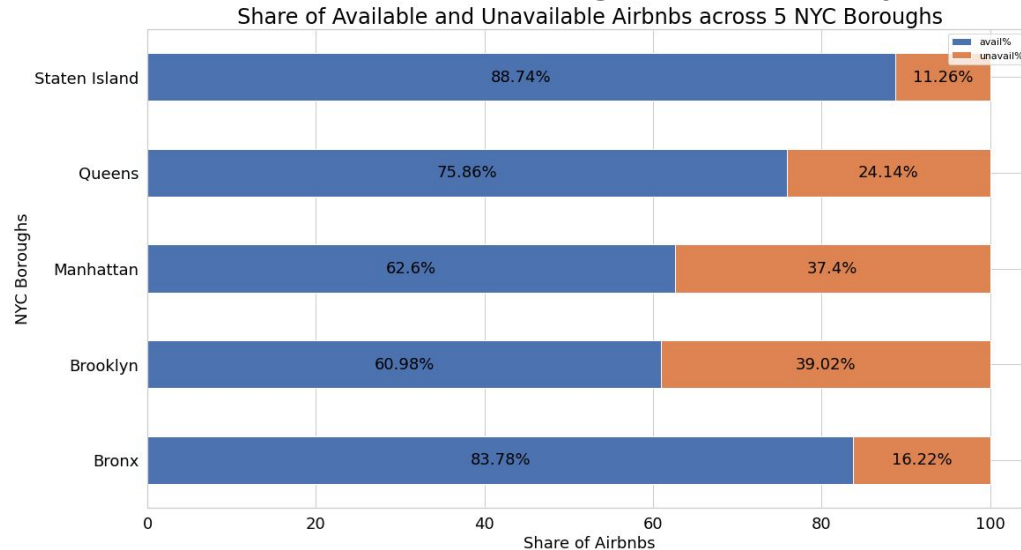


- Basis “Busyness over entire period”, Silver Lake neighbourhood of Staten Island borough is the busiest neighbourhood
- Basis “Busyness per month”, East Elmhurst neighbourhood of Queens borough is the busiest neighbourhood
- Most of the top 10 busiest neighbourhoods belongs to Queens and Staten Island



# 7. Available Vs Unavailable Airbnb

- **Assumption:** 0-day availability implies that Airbnb is unavailable for booking during the next 365 days, and any other positive value implies that the Airbnb is available for booking in next 365 days
- The share of Unavailable Airbnb is largest in Brooklyn and is smallest in Staten Island



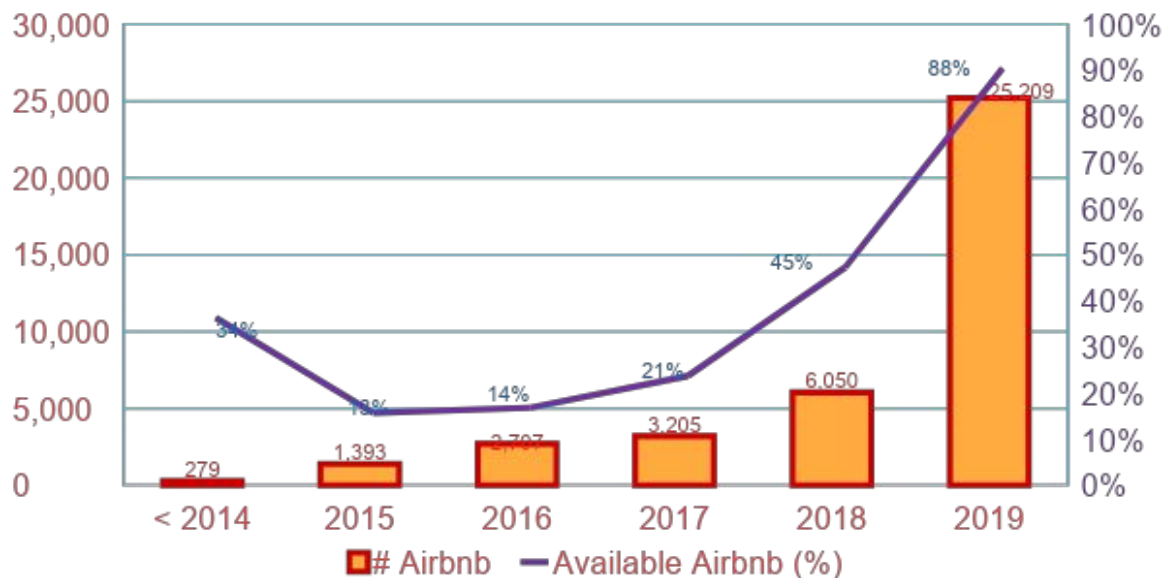
- Basis above plot, Airbnb may investigate the reasons for unavailability and can further devise strategies to increase hosts' engagement on its platform

## 8. Findings basis year of Last Review

- **Assumption:** A given Airbnb was considered active during the year of last review. Basis this, plotted year-wise count of followings:
  1. # Airbnbs,
  2. Percentage of Airbnbs which are available for booking during the next 365 days
- Dates of last review range from 28-3-2011 to 08-07-2019. So, there are Airbnbs which were not active for a very long time. But a portion of these airbnbs have now become available for booking in next 365 days.

Year	# Airbnb	Available Airbnb's Share (%)
2011	7	57 %
2012	25	60 %
2013	48	48 %
2014	199	26 %
2015	1393	13 %
2016	2707	14 %
2017	3205	21 %
2018	6050	45 %
2019	25209	88 %

Year-wise distribution of Airbnb &amp; Available Airbnb



- Share of Available Airbnb out of last reviewed Airbnbs in a year have increased significantly every year
- Around 52% Airbnb (25,209) were last reviewed in year 2019 and majority of these Airbnb (~88%) are available for booking during the next 365 days



# Agenda

**A. Introduction**

**B. Data Preparation & Cleaning**

**C. Data Exploration and Visualization**

**D. Conclusion**

# Conclusions

- Manhattan & Brooklyn have the highest share of Airbnb and hosts, while Staten Island has the lowest
- Queens & Staten Island are the busiest basis average traffic per Airbnb
- Most Airbnbs are either "Entire home/apt" or "Private room" type
- Basis average number\_of\_reviews, 'Private room' type Airbnb is the busiest property type, followed by 'Entire Home' type and lastly 'shared room' type
- Most Airbnbs (75%) require  $\leq 5$  nights of minimum stay. So, short term stays are more popular
- "Entire home/apt" type Airbnb in Manhattan is the most expensive Airbnb; while "Shared room" type Airbnb in Staten Island is the most affordable

**END OF PRESENTATION  
THANKS**