Storeytelling Case Study: Airbnb

Presentation-2

(To Head of Acquisitions and Operations, NYC and Head of User Experience, NYC)

Submitted By:

Ranjith
Venkatesh
Rachit

Agenda

- Objective

- Background

- Key Findings

- Recommendations

- Appendix:
  - Data Source
  - Data Assumptions
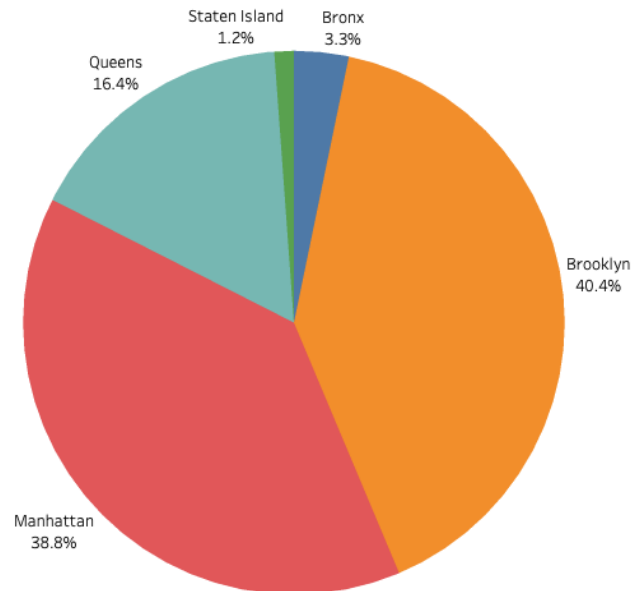 - Data Methodology

Objective

- To understand customer preferences with respect to properties
- Know which type of properties to acquire and where
- Optimize traction in various properties

Background

- For the past few months, Airbnb has seen a major decline in revenue.

- Now COVID restrictions have started lifting and people have started to travel more

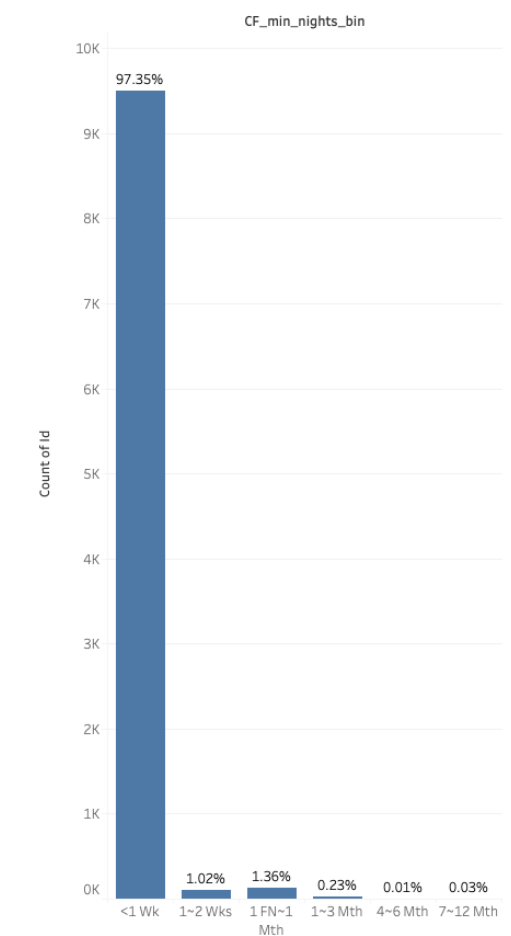- Airbnb wants to make sure that it is fully prepared for this change.

Key Finding 1:
# Customer preference: Room Type, Neighbourhood Group, Stay duration, Availability
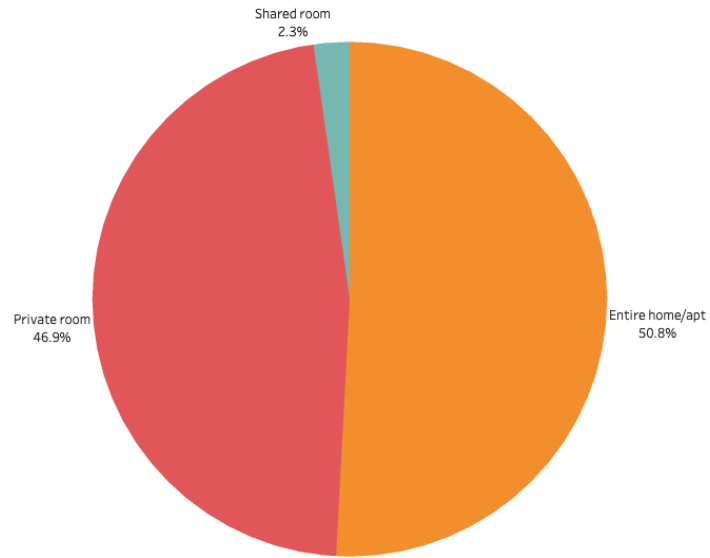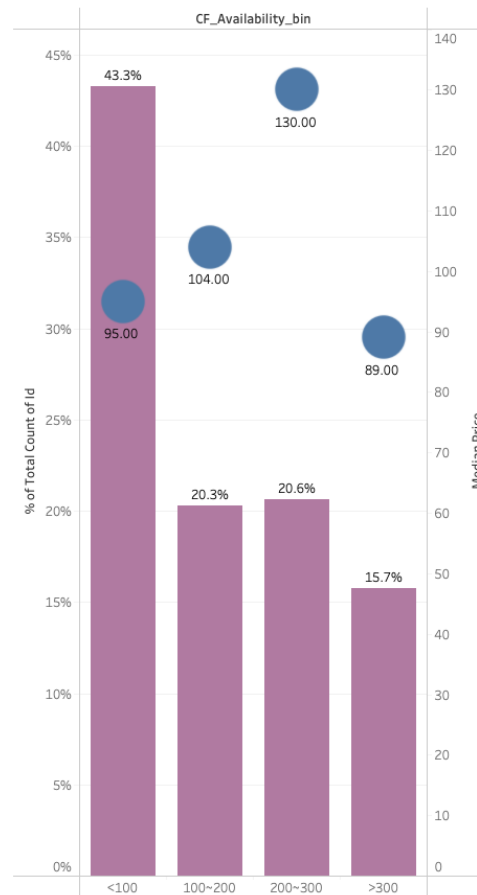
- Entire home/Apt is more popular, followed by Private room
- Brooklyn has most number of popular properties, followed by Manhattan
- Most of the customers prefer <1 week stay
- Popular properties which are available for 200~300 days are able to get the highest median price of 130

Key Finding 2:
Preferred Properties in popular Neighbourhoods at desired prices

Good Popularity

Neighbourhood Group / Room Type
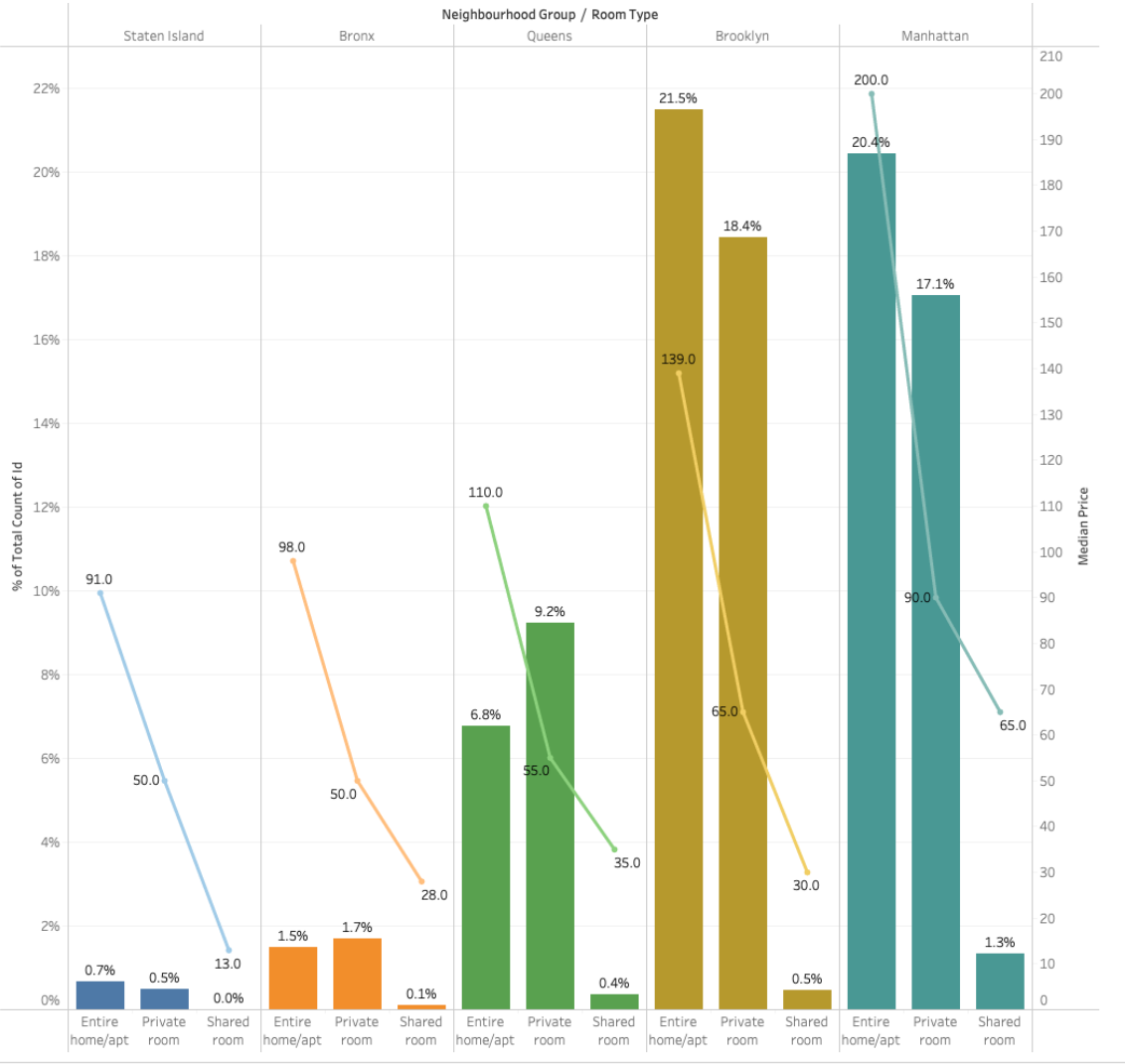
- Add more properties basis room type, Neighbourhood and median price, as mentioned below:

|	Manhattan	Brooklyn	Queens	Bronx	Staten Island
Entire Home/ Apt		139	110	98	91
Private Room	90		55	50	50
Shared Room	65				
Neighbourhood	Theatre District	Canarsie, East New York, Flashbush	East Elmhurst, Jamaica and Richmond Hill	Claremont Village, Belmont	Mariners Harbor, Grant City and Stapleton

- Review prices of poor popularity properties:
  Shared room prices are very high for poor popularity properties in Bronx, Staten Island
  Prices are in general high for poor popularity properties in Staten Island

- Median price need not be changed in Manhattan, Brooklyn, Queens

Key Finding 3:
Property Availability in Popular Neighborhoods

Availability by Region

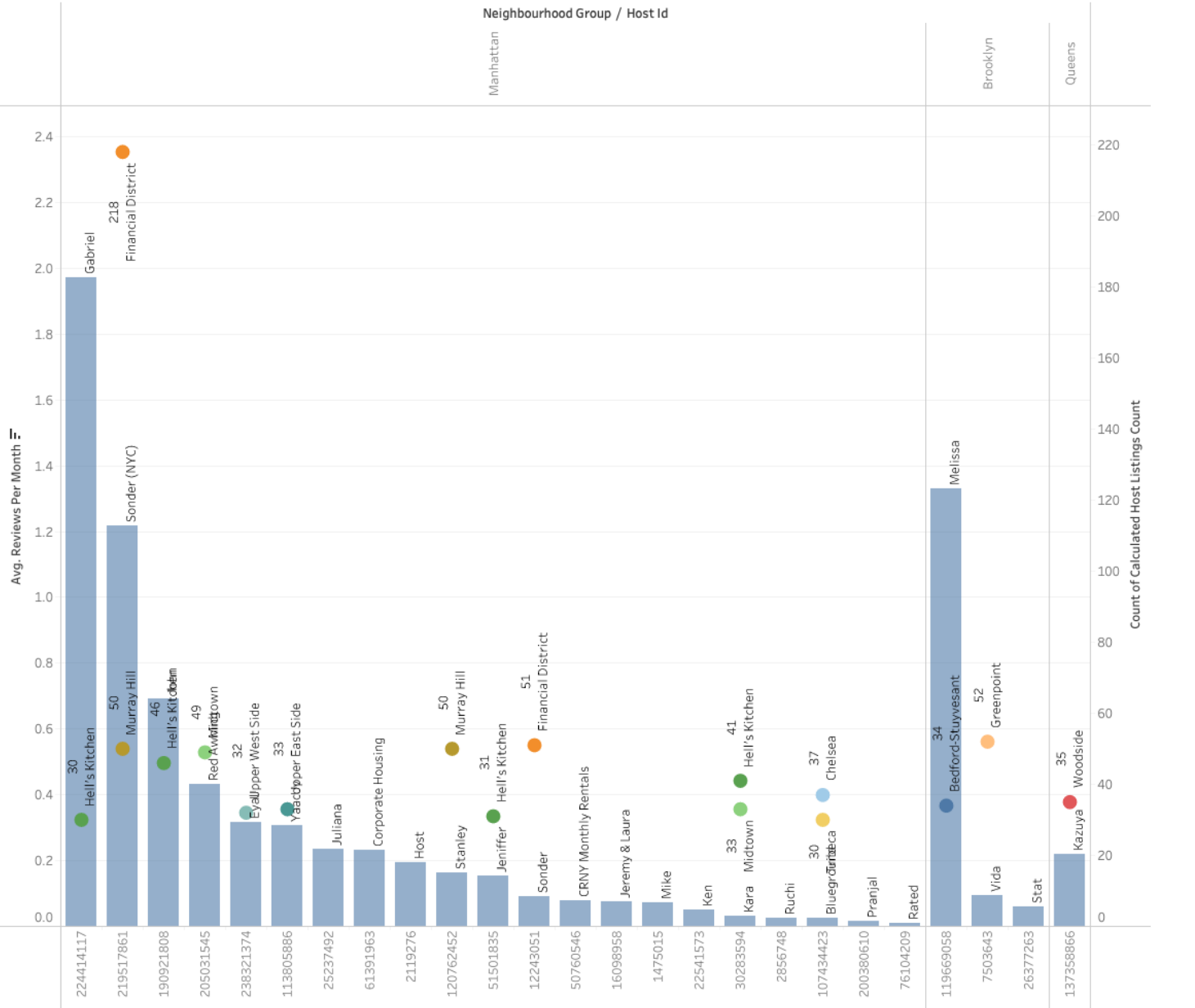| CF_Availabilit...	Bronx	Brooklyn	Manhattan	Queens	Staten Isla..
>300	0.63%	5.64%	7.46%	2.70%	0.27%
<100	0.98%	26.82%	26.95%	5.80%	0.26%
200~300	0.22%	4.13%	4.93%	1.23%	0.13%
100~200	0.39%	4.51%	4.97%	1.85%	0.11%

Availability of popular properties
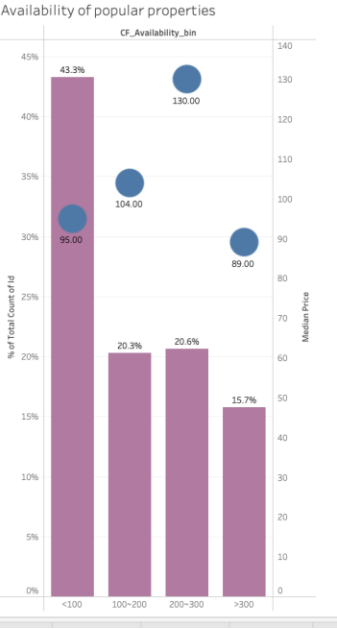
Availability in popular Neighbourhoods
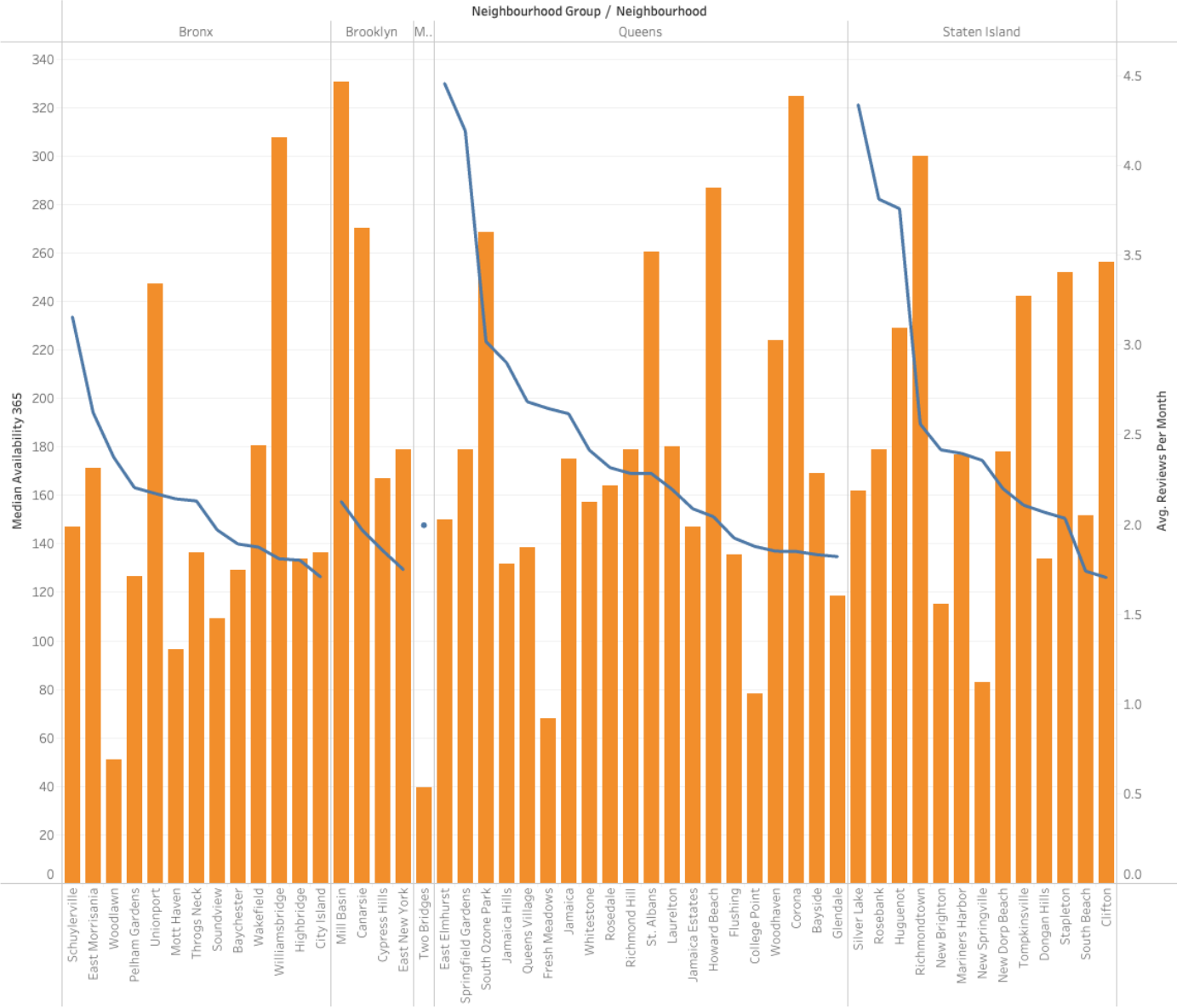
- 61 % of all the properties have availability less than 100 days and most of them are located in Manhattan and Brooklyn

- 43.3% of the Popular properties are available for less than 100days. Median price of 130 is the highest for availability of 200~300 days

- Following neighborhoods should make their properties available for more number of days in the year considering the demand in those areas
  - Bronx: Woodlawn, Mott Haven
  - Manhattan: Two Bridges
  - Queens: East Elmhurst, Springfields Garden Jamaica Hills, Fresh Meadows
  - Staten Island: Silver Lake, Rosebank, New Brighton, New Springville

Key Finding 4:
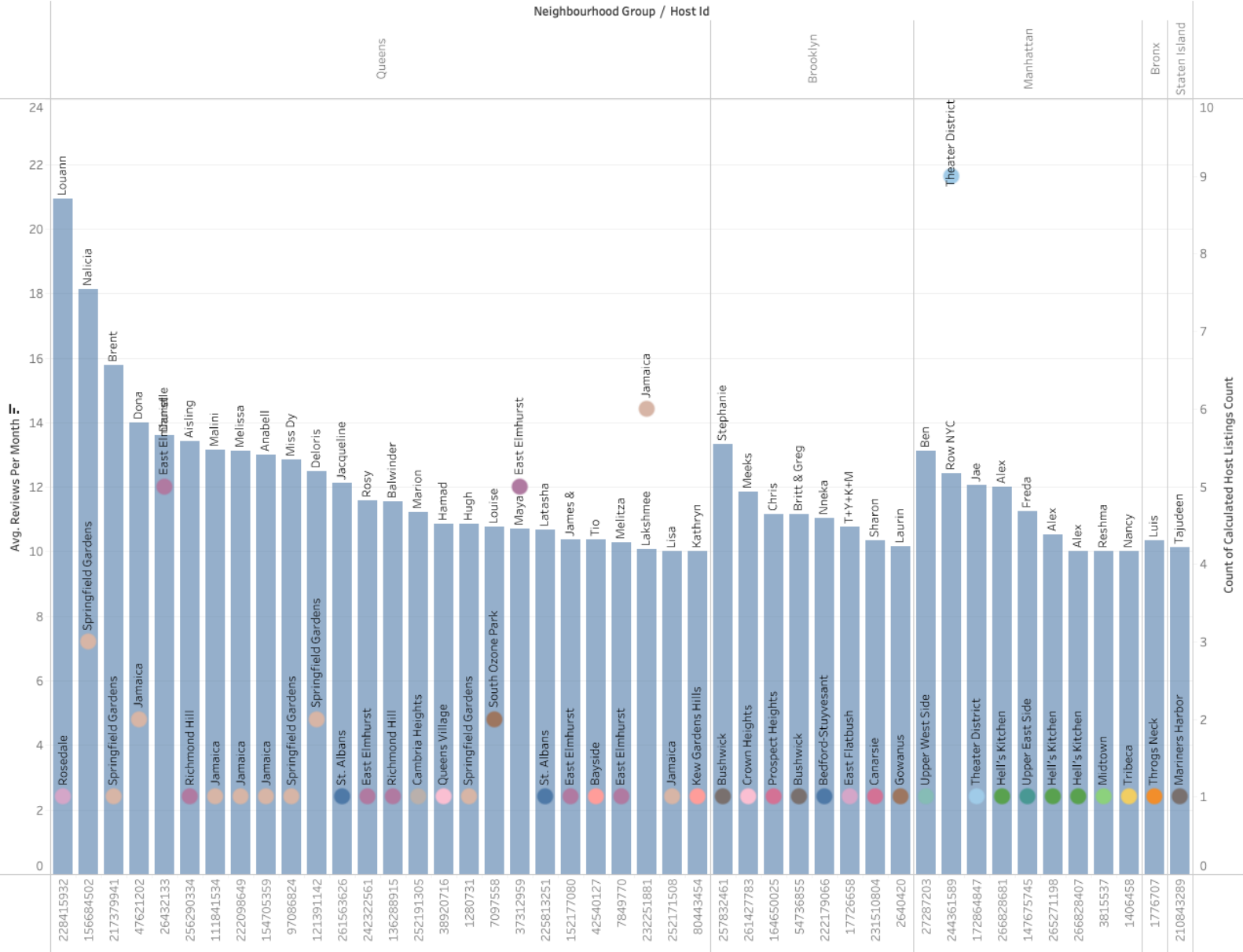# Focus on Hosts having more than 30 properties

Hosts having more than 30 properties

- Most of the hosts with >30 properties are in Manhattan.
- Brooklyn has 3 and Queens has only one such host.
- No such hosts in Bronx and Staten Island

- Most of such hosts have less than 0.5 reviews per month
- More focus to be given on hosts having large number properties to improve traction
- Hosts like Blueground, Kara,Pranjal, Rated to be focused more for improvement

Key Finding 5:
Benchmark and Encourage Hosts having best reviews

Hosts with having >10 reviews per month

- Top hosts be benchmarked for other properties:
  Queens: Louann,Nalicia,Brent
  Brooklyn: Staphanie, Meeks,Chris
  Manhattan:Ben, Pow NYC,Jae
  Bronx: Luis
  Staten Island:Tajudeen

- 80% of most popular hosts have only one property. They should be incentivised and encouraged to open more properties.

- Following Neighbourhoods have more popular hosts:
  Queens: Springfields Garden, Jamaica, Richmond Hill
  Brooklyn: Bushwick
  Manhattan: Theatre District, Hells' Kitchen

  These localities can be considered for addition of hosts

Recommendations

- Acquire properties as mentioned in the table below:

	Manhattan	Brooklyn	Queens	Bronx	Staten Island
Entire Home/ Apt (Price)		139	110	98	91
Private Room (Price)	90		55	50	50
Shared Room (Price)	65				
Neighbourhood	Theatre District	Canarsie, East New York, Flashbush	East Elmhurst, Jamaica and Richmond Hill	Claremont Village, Belmont	Mariners Harbor, Grant City and Stapleton

- Improve availability at the following locations:
  - Bronx: Woodlawn, Mott Haven
  - Manhattan: Two Bridges
  - Queens: East Elmhurst, Springfields Garden Jamaica Hills, Fresh Meadows
  - Staten Island: Silver Lake, Rosebank, New Brighton, New Springville

- Get more traction for high number of less popular properties
  - More focus to be given on hosts having large number properties
  - Min nights to be reduced to below 1 week. Availability to be targeted between 200~300 days.
  - Benchmark top hosts for other properties. Incentivise and encourage popular hosts to open more properties

Appendix 1: Data Source

Dataset provided by Airbnb comprising of following fields:

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Appendix 2: Data Assumptions

- Number of reviews considered to be a metric of popularity of a property. It was assumed to be an indicator of customer preference.

- >2 reviews per month was used as a metric of good popularity while <0.5 reviews per month was considered to be poor popularity

- Null values in "name" and "host_name" fields were ignored and assumed to have no significance

- Missing values in "reviews_per_month" were assumed to correspond to 0 reviews.

Appendix 3: Data Methodology

Steps followed

- Data sourcing to Excel

- Missing value treatment

- Outlier treatment

- Import in Tableau

- Binning of continuous variables

- Check current distribution of properties by Room Type , Location and Price etc using Tableau Visualisation

- Check relation of customer preference (reviews) with respect to Room type, location , Price etc using Tableau Visualisation

- Analyse the data and gather insights. Use review as a measure of preference of customers

- Make recommendations

Data sourcing to excel

- The dataset provided as sourced to Excel
- Following fields are present in the data

Column	Description
id	listing ID
name	name of the listing
host_id	host ID
host_name	name of the host
neighbourhood_group	location
neighbourhood	area
latitude	latitude coordinates
longitude	longitude coordinates
room_type	listing space type
price	
minimum_nights	amount of nights minimum
number_of_reviews	number of reviews
last_review	latest review
reviews_per_month	number of reviews per month
calculated_host_listings_count	amount of listing per host
availability_365	number of days when listing is available for booking

Check missing values, outliers and treat them

- The dataset has 48895 records
- Some columns have missing records as shown below:

	name	host_name	last_review	reviews_per_month
No of Missing record	16	21	10052	10052
% of missing records	0%	0%	21%	21%

- "name " and "host_name" have very less missing values. Also their respective ids don't have any missing values. Hence no missing value treatment is done.
- All "last_review" and "reviews_per_month" missing values correspond to the records where there are no reviews.
- Hence missing values in "reviews_per_month" are imputed with "0" and "last_review" missing values are replaced with "No_reviews".
- Price column has 11 records having 0 value. Since the number of records are very less, they have been deleted.

Import in Tableau and Binning of continuous variables

- Import the data to Tableau for visualisation

- Price ranges from 1~10000. Appropriate bins were created. Only 0.2% properties have price >2000.

Price vs ID
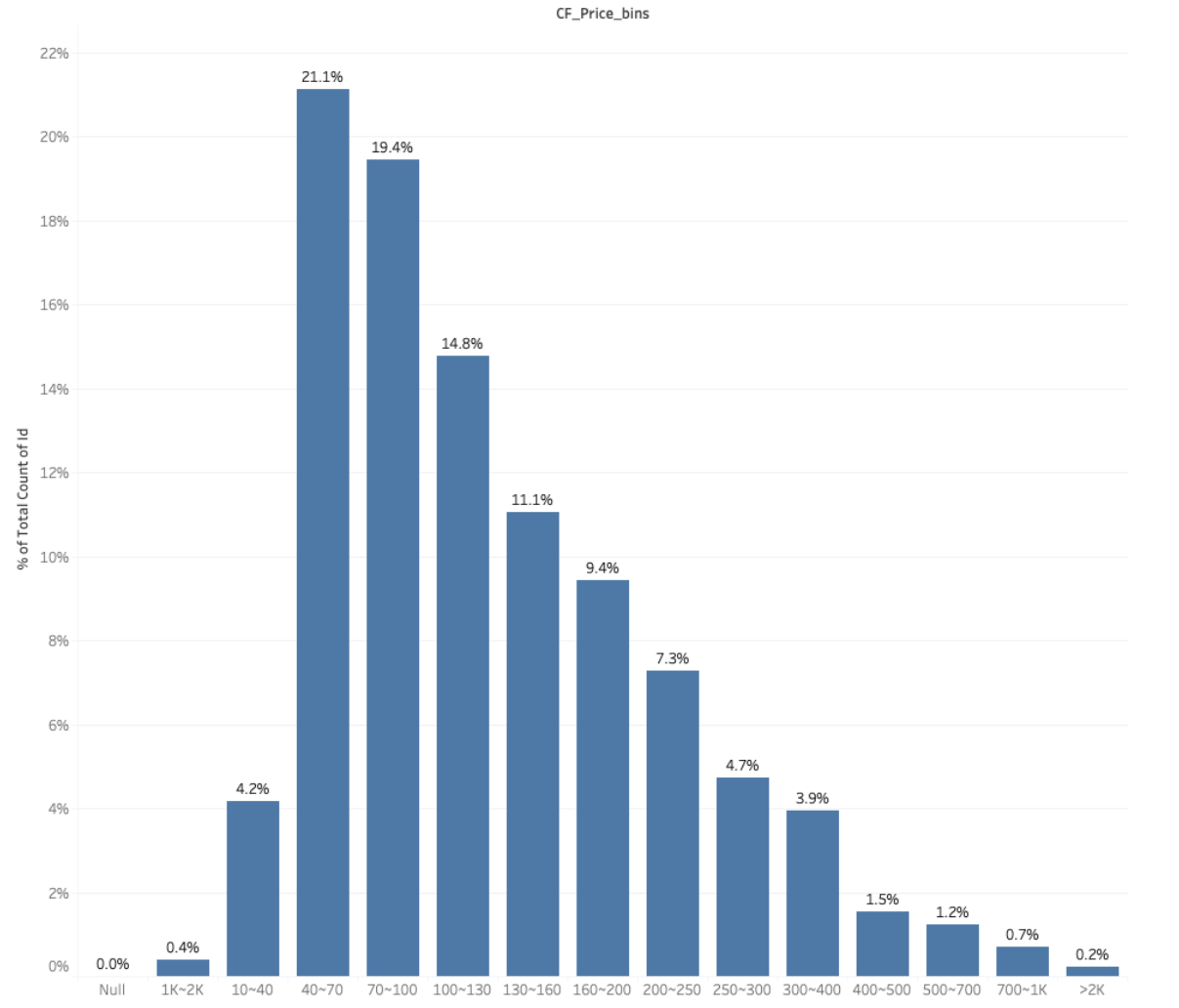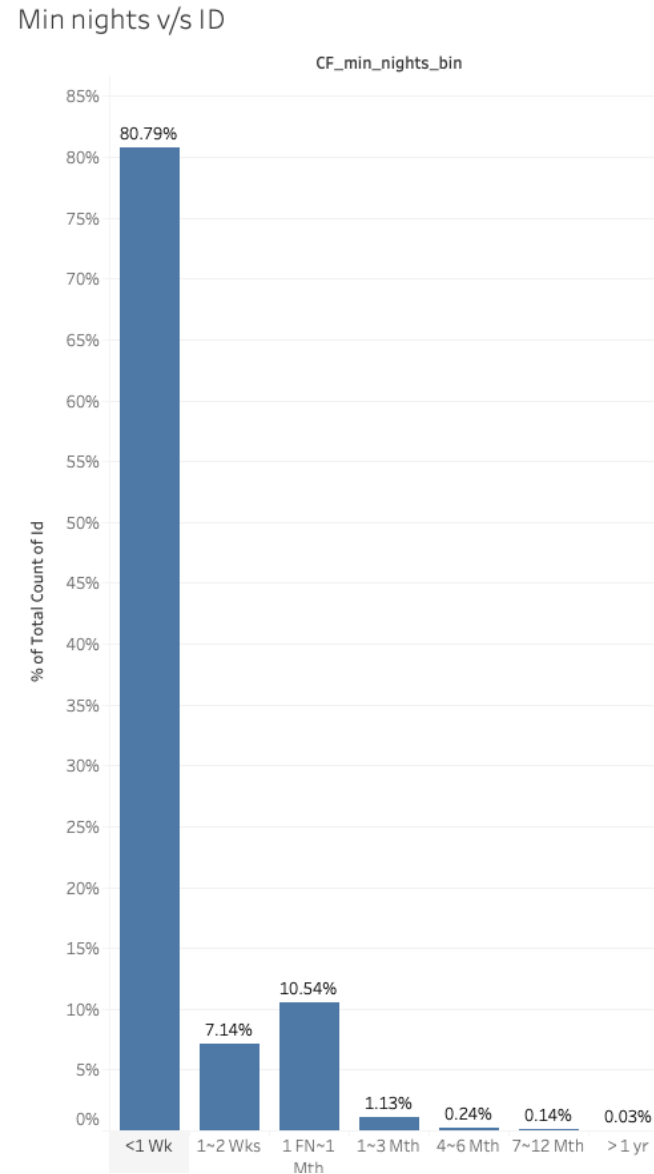
Binning of continuous variables: Availability 365

- For Availability 365 (Number of days) of rooms 4 bins were created <100,100-200,200-300,>300. 16.7% properties have >300 days availability.60.8% properties have <100 days availability

Availability vs ID
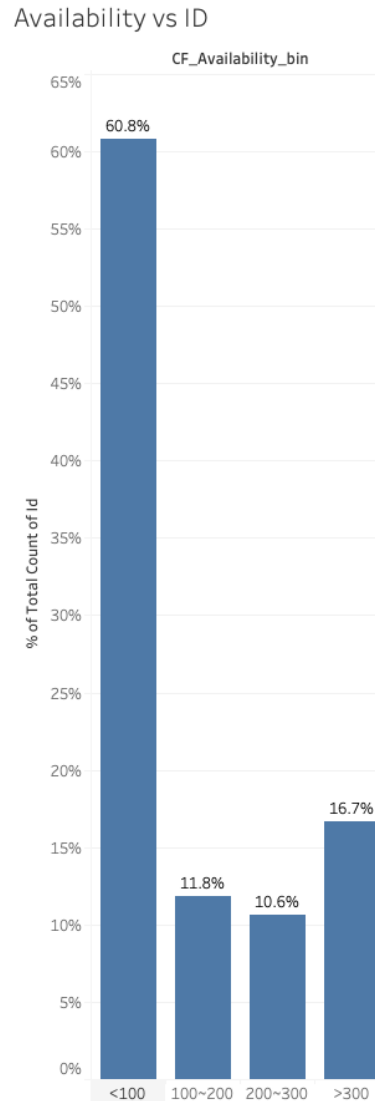
Binning of continuous variables: Minimum nights

- Minimum nights ranges from 1~ 1250. Total 7 bins were created. >1 Yr is only 0.03%. 80.79% properties have Min Nights <1 week

Min nights v/s ID

CF_min_nights_bin

Binning of continuous variables: Reviews Per Month

- 4 Bins were created for Reviews Per Month. 20.6% properties have No Reviews. Only 3% properties have >5 reviews per month.

- Medium and High review properties were considered to fall in Good popularity category while Poor review properties were considered to be not popular. Filtering as per this criterion was used in subsequent analysis

- Total 20% properties fall in Good popularity category
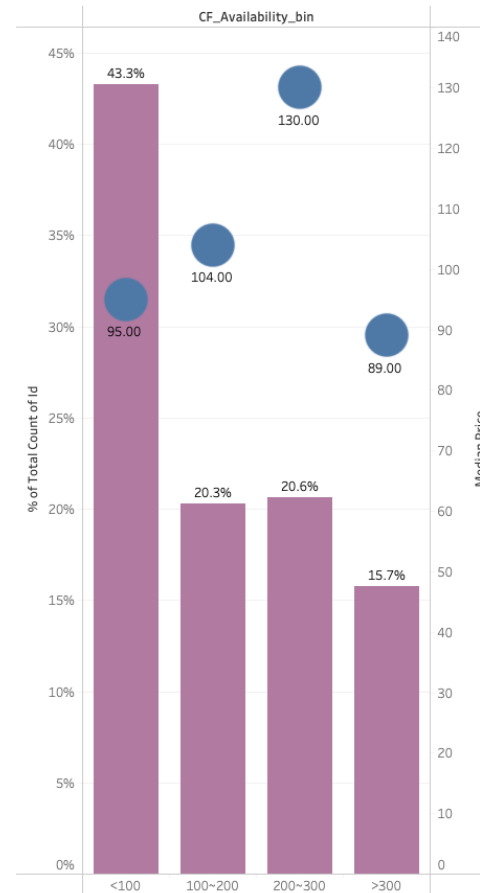

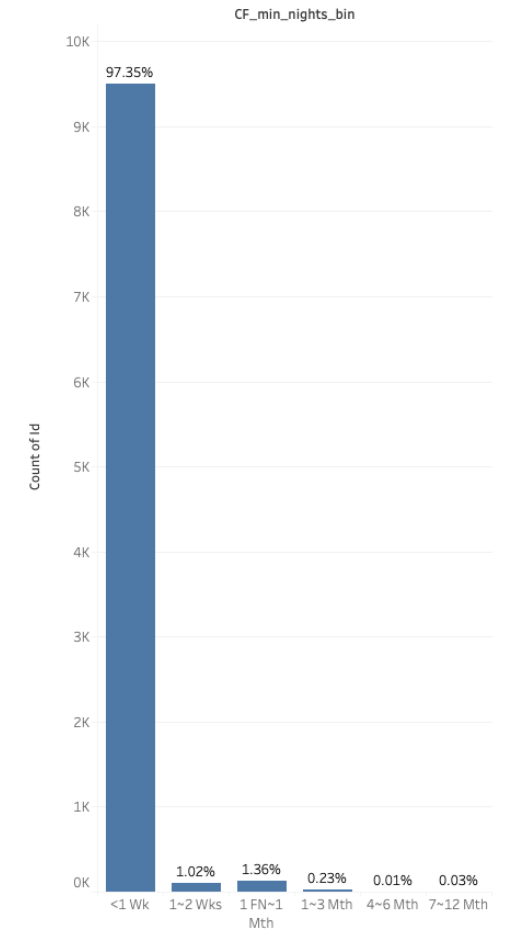Distribution of review per month
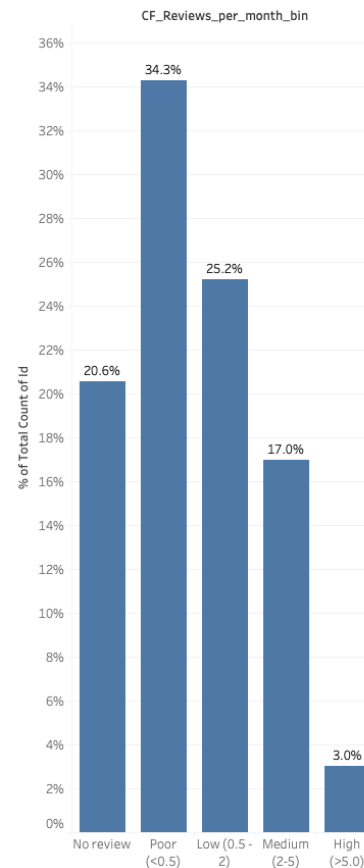
Analysis 1:
Customer preference: Room Type, Neighbourhood Group, Stay duration, Availability
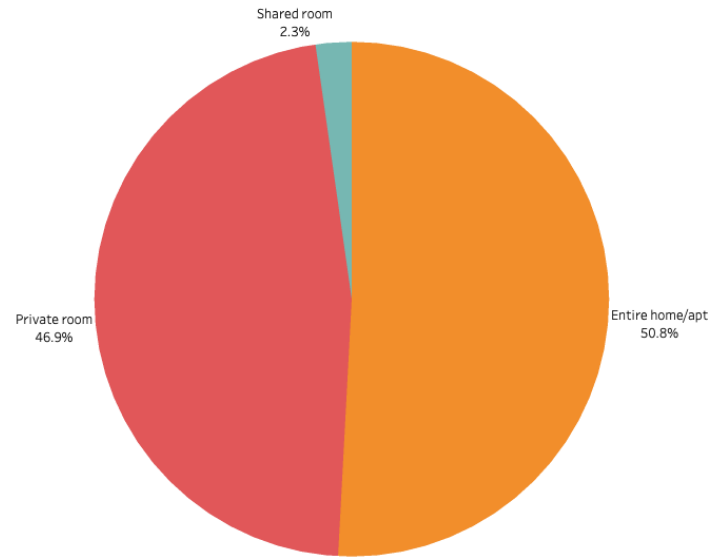

- Filtering in Tableau was done for Good Popularity properties for creating all these visualisations
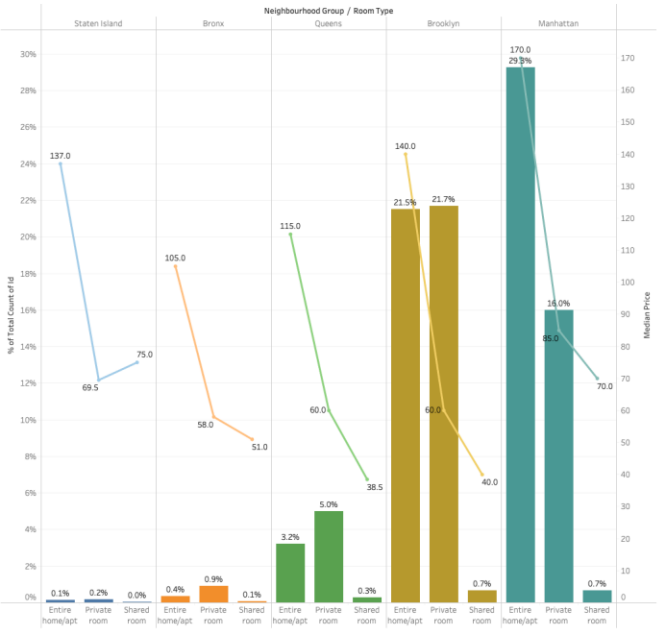
Analysis 2:
Preferred Properties in popular Neighbourhoods at desired prices

- Good and Poor popularity properties were visualised using dual axis charts.
- The properties were mapped wrt Neighbourhood Group, Room type and Median price
- Further, for each neighbourhood group property types were mapped wrt popularity and median price
- Finally number of properties in top neighbourhood were checked with respect to their popularity.
- Gaps in number of properties were identified and high potential localities were noted down
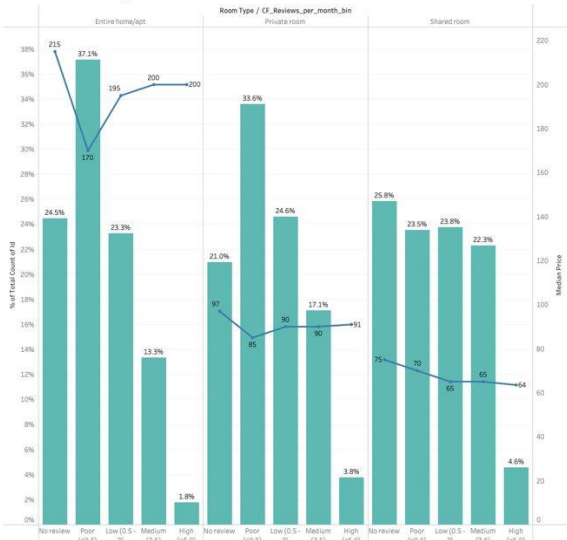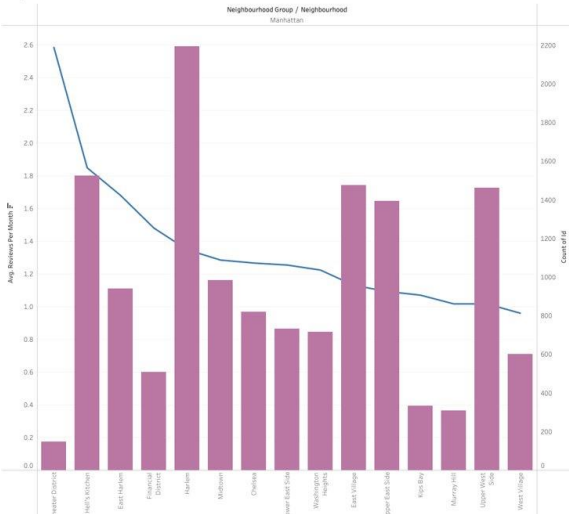

Good Popularity
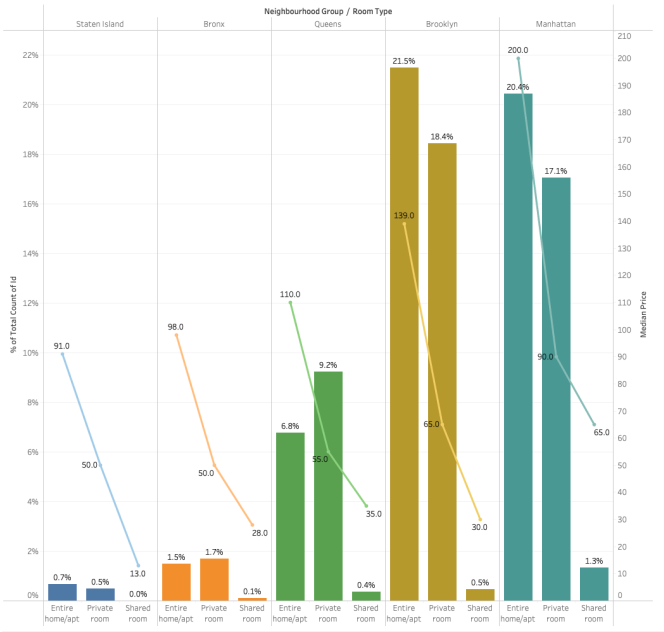
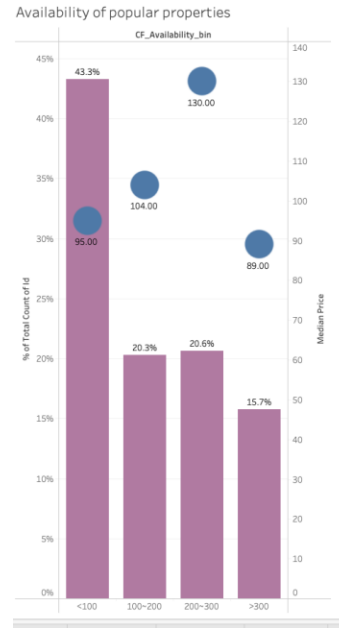Poor Popularity

Distribution of Room Type-Manhattan
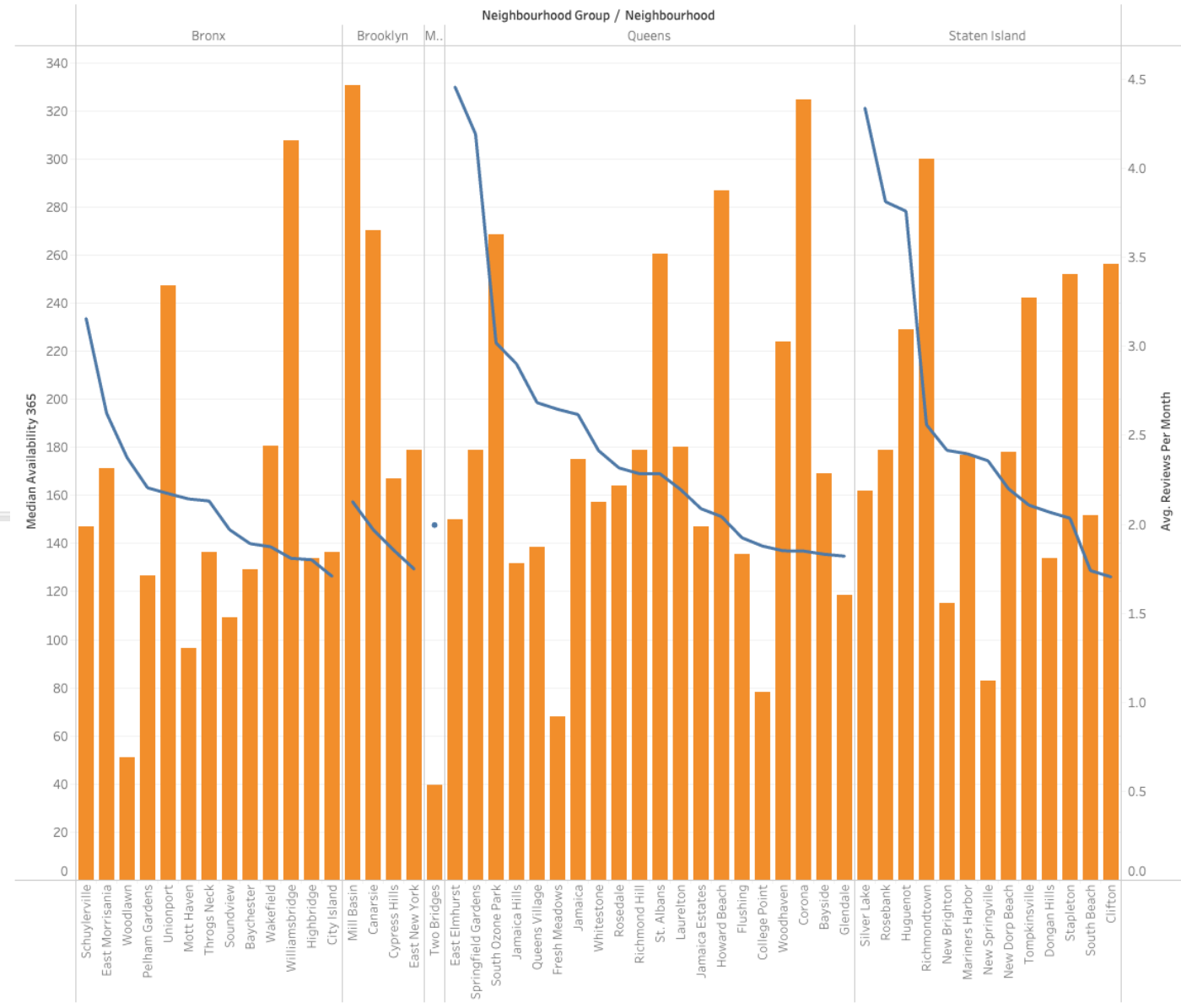

Prop distribution and Reviews Per month

Analysis 3:
Property Availability in Popular Neighborhoods

Availability by Region

CF_Availabilit..	Bronx	Brooklyn	Manhattan	Queens	Staten Isla..
>300	0.63%	5.64%	7.46%	2.70%	0.27%
<100	0.98%	26.82%	26.95%	5.80%	0.26%
200~300	0.22%	4.13%	4.93%	1.23%	0.13%
100~200	0.39%	4.51%	4.97%	1.85%	0.11%

Availability of popular properties

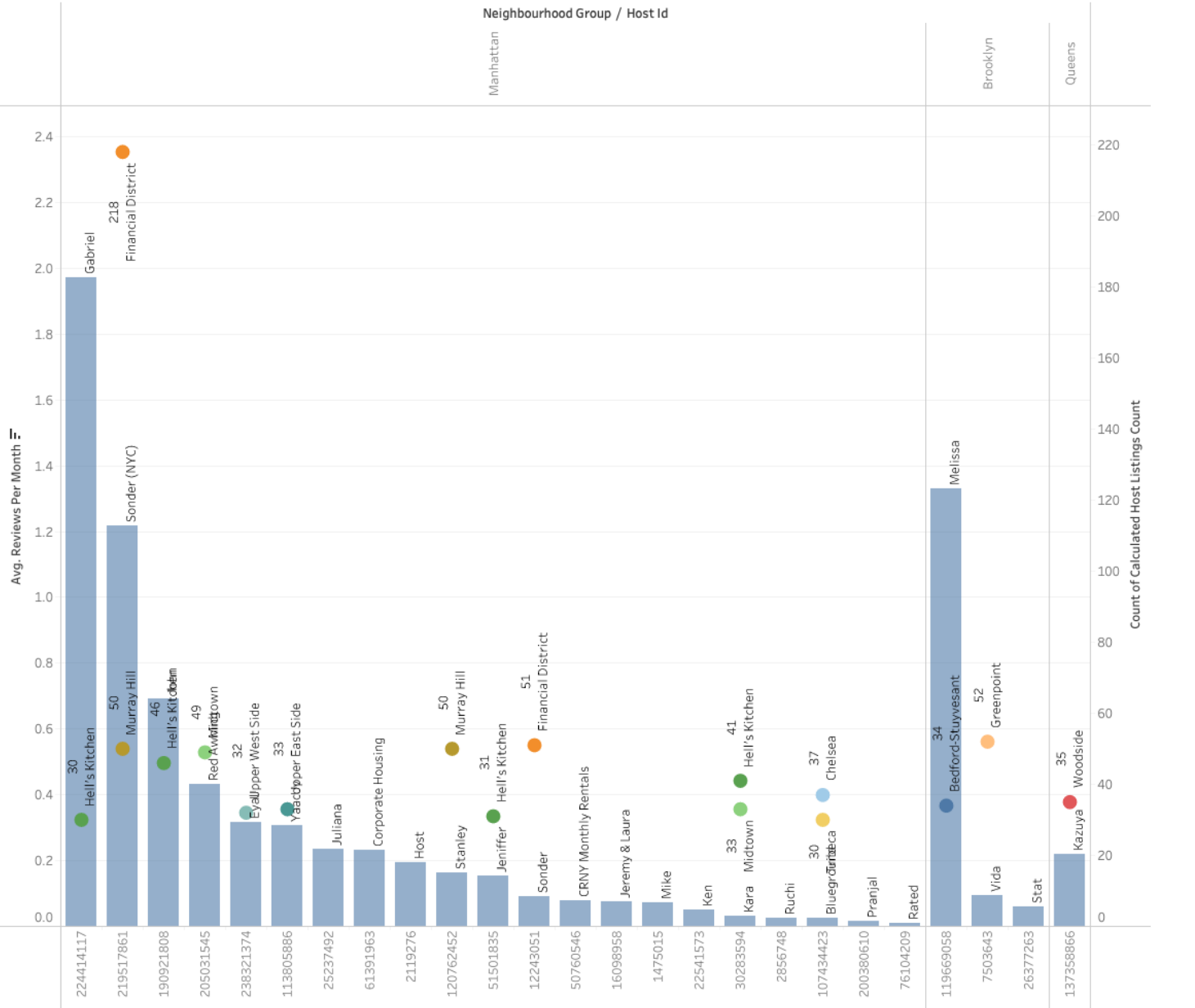Availability in popular Neighbourhoods

- Region-wise availability was checked wrt the bins created earlier
- Another visualization was created to check Good popularity properties wrt availability and median price
- Using dual axis chart a visualization was created to map neighborhoods wrt availability and average monthly review and gaps were identified

Analysis 4:
Check popularity of Hosts having more than 30 properties

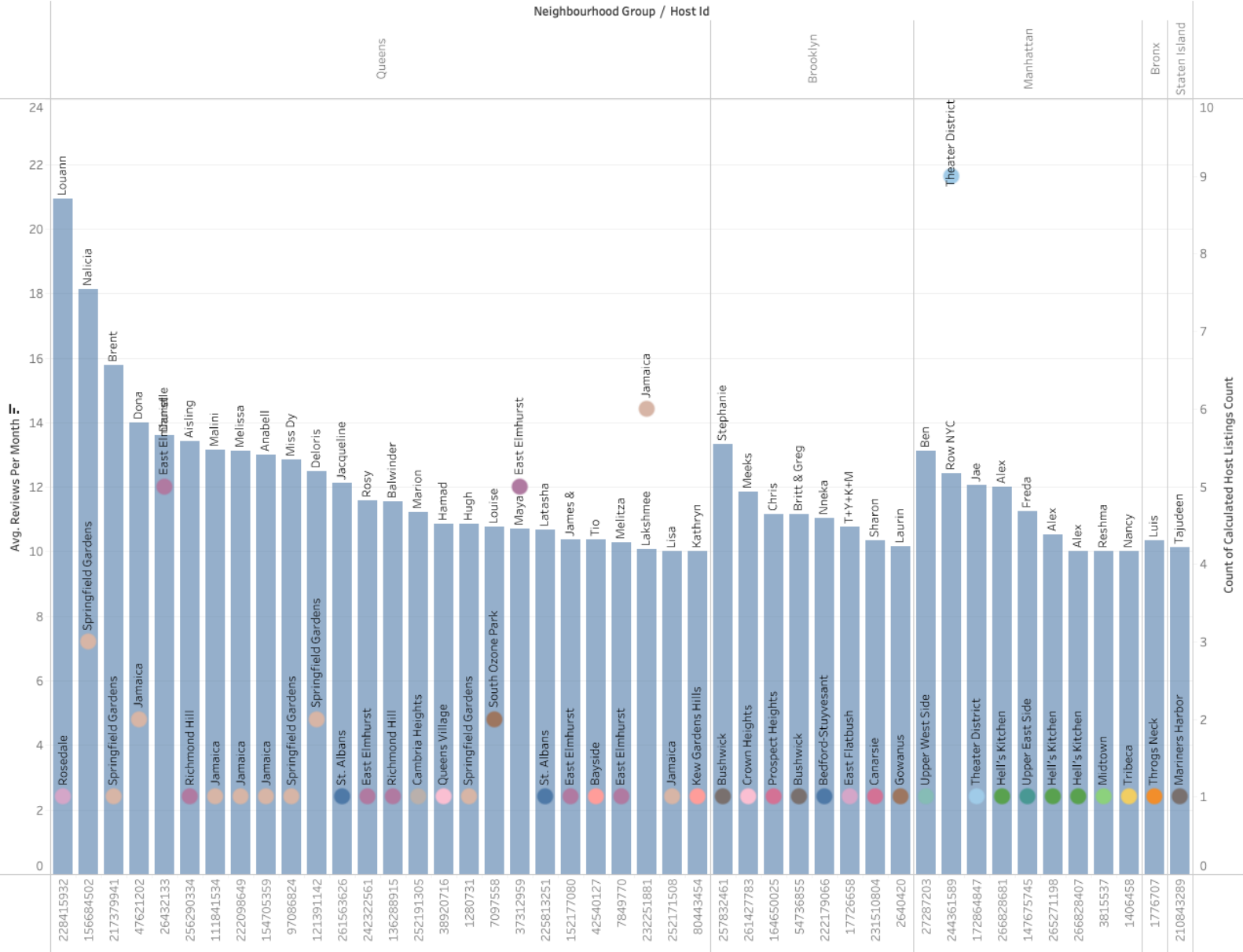Hosts having more than 30 properties

- Dual axis chart was made to check the popularity of hosts having more than 30 properties.

Analysis 5:
Find out the Hosts having best reviews and where they are located

Hosts with having >10 reviews per month

Another dual axis chart was made to find out hosts having best reviews.

Neighbourhoods belonging to these hosts were also highlighted

Make Recommendations

- Using the analysis preferred room type , median price and neighbourhoods were summarised as recommended properties to be acquired:

	Manhattan	Brooklyn	Queens	Bronx	Staten Island
Entire Home/ Apt (Price)		139	110	98	91
Private Room (Price)	90		55	50	50
Shared Room (Price)	65				
Neighbourhood	Theatre District	Canarsie, East New York, Flashbush	East Elmhurst, Jamaica and Richmond Hill	Claremont Village, Belmont	Mariners Harbor, Grant City and Stapleton

- Basis the analysis locations were identified where improvement in availability will help in boosting the business. These locations were recommended

- Recommendations were made to get more traction for high number of less popular properties basis
  - Analysis of hosts having large number of properties
  - Analysis of popularity of Min nights stay
  - Analysis of popularity basis Availability .
  - Analysis of hosts having Good popularity