FLH CASE STUDY

Business Context

“Feel Like Home”(aka FLH) is a home rental company like Oyo. For the past few months, because of covid, FLH has seen a major decline in revenue. Now that the restrictions have started lifting and people have started to travel more, the organization wants to make sure that it is fully prepared.

Problem Statement

The different leaders at FLH want to understand some important insights based on various attributes in the dataset to maximize revenue during this period of increasing travel:

- What are the neighbourhoods that FLH needs to target?
- What are the pricing ranges preferred by FLH customers?
- What are the types of properties that are most successful and least successful for FLH?
- What are customers looking for most in their stays with FLH?
- What attributes do the hosts have that FLH should target to add to their service to increase revenue?
- How to get unpopular properties more traction?

Data Availability

- This dataset has around 49,000 observations in it with 16 columns and it is a mix between categorical and numeric values.
- **Categorical variable(Nominal):** Neighbourhood group, Neighbourhood and Room type
- **Numerical Variable(Continous):** Price
- **Numerical Variable(Descrete):** Minimum Nights, Number of reviews, Last review, Reviews per month, Calculated host listings count, Availability 365
- **Location variable:** Latitude, Longitude
- It all needed information to find out more about hosts, geographical availability, necessary metrics to make predictions and draw conclusions.

Data Preprocessing

- Data Preparation.ipynb

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   id                              48895 non-null  int64
 1   name                            48879 non-null  object
 2   host_id                         48895 non-null  int64
 3   host_name                       48874 non-null  object
 4   neighbourhood_group             48895 non-null  object
 5   neighbourhood                   48895 non-null  object
 6   latitude                        48895 non-null  float64
 7   longitude                       48895 non-null  float64
 8   room_type                       48895 non-null  object
 9   price                           48895 non-null  int64
 10  minimum_nights                  48895 non-null  int64
 11  number_of_reviews               48895 non-null  int64
 12  last_review                     38843 non-null  object
 13  reviews_per_month               38843 non-null  float64
 14  calculated_host_listings_count  48895 non-null  int64
 15  availability_365                48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

Missing Values

	n_rows	n_cols	null_vals	dupl_vals
0	48895	16	20141	0

Missing Value treatment

```
#replacing all NaN values in 'reviews_per_month' with 0
data.fillna({'reviews_per_month':0}, inplace=True)
#examing changes
data.reviews_per_month.isnull().sum()
```

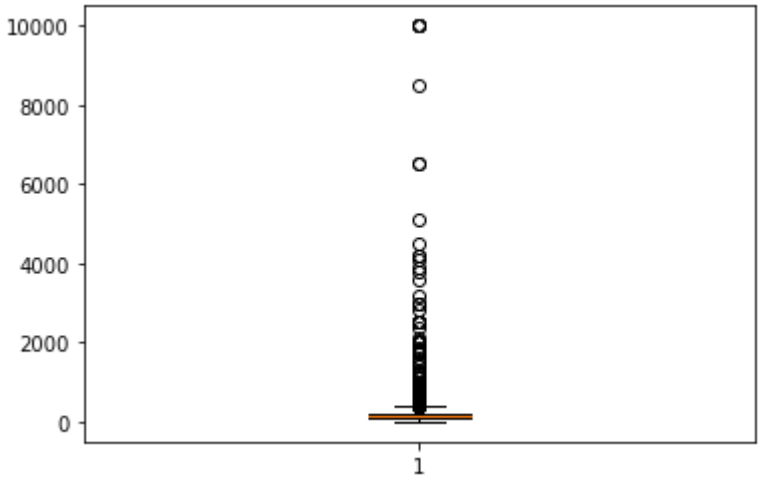
```
0
```

```
#dropping columns that are not significant for our data exploration and analysis
data.drop(['last_review','name'], axis=1, inplace=True)
#examing the changes
data.head(3)
```

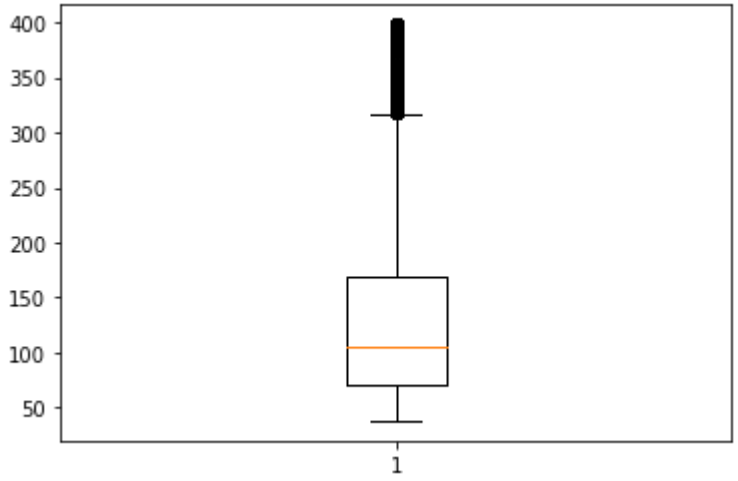
	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1

Outliers

- Outliers for continuous numerical variable are checked and eliminated



Before Treatment



After Treatment

Assumption

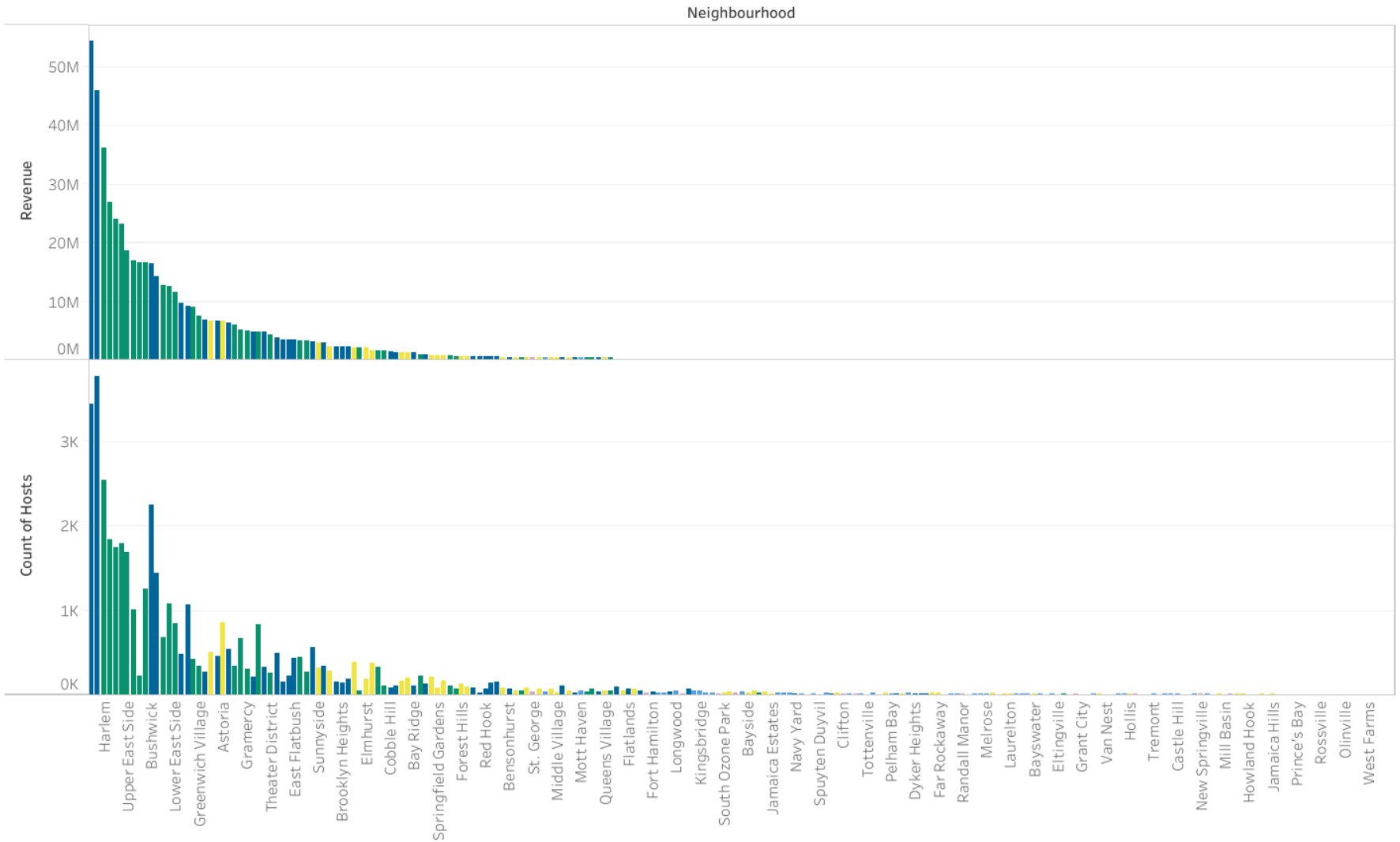
- Assumption:
  - Number of Reviews = Number of Customer to That particular host
  - Price is for sing night stay
- Here target variable is not explicitly given, so to measure the performance of host I have added new attribute called by Revenue by multiplying Price and Number of customer and their minimum night stays

```
:
data2["revenue"]=data2["price"]*data2["number_of_reviews"]*data2["minimum_nights"]
data2
```

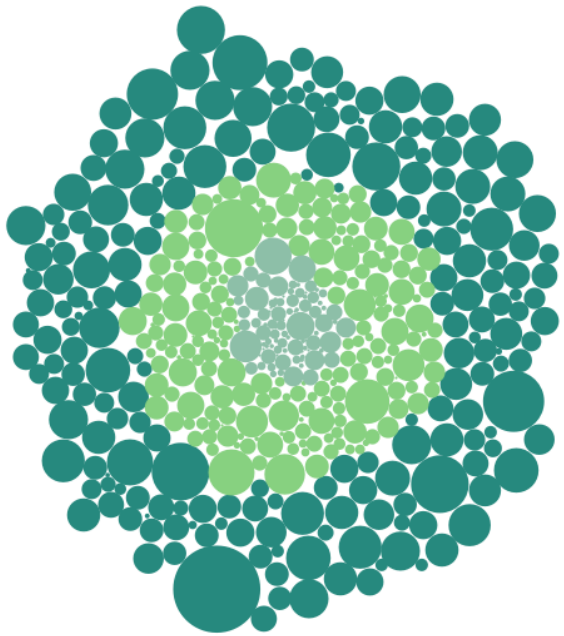
Data Analysis

Complete Data Analysis

NEIGHBOURHOOD TO BE CONCENTRATED BY FLH



Room Type
- ✔ Entire home/apt
- ✔ Private room
- ✔ Shared room



Observation
Neighborhood with small circle needs to be targeted

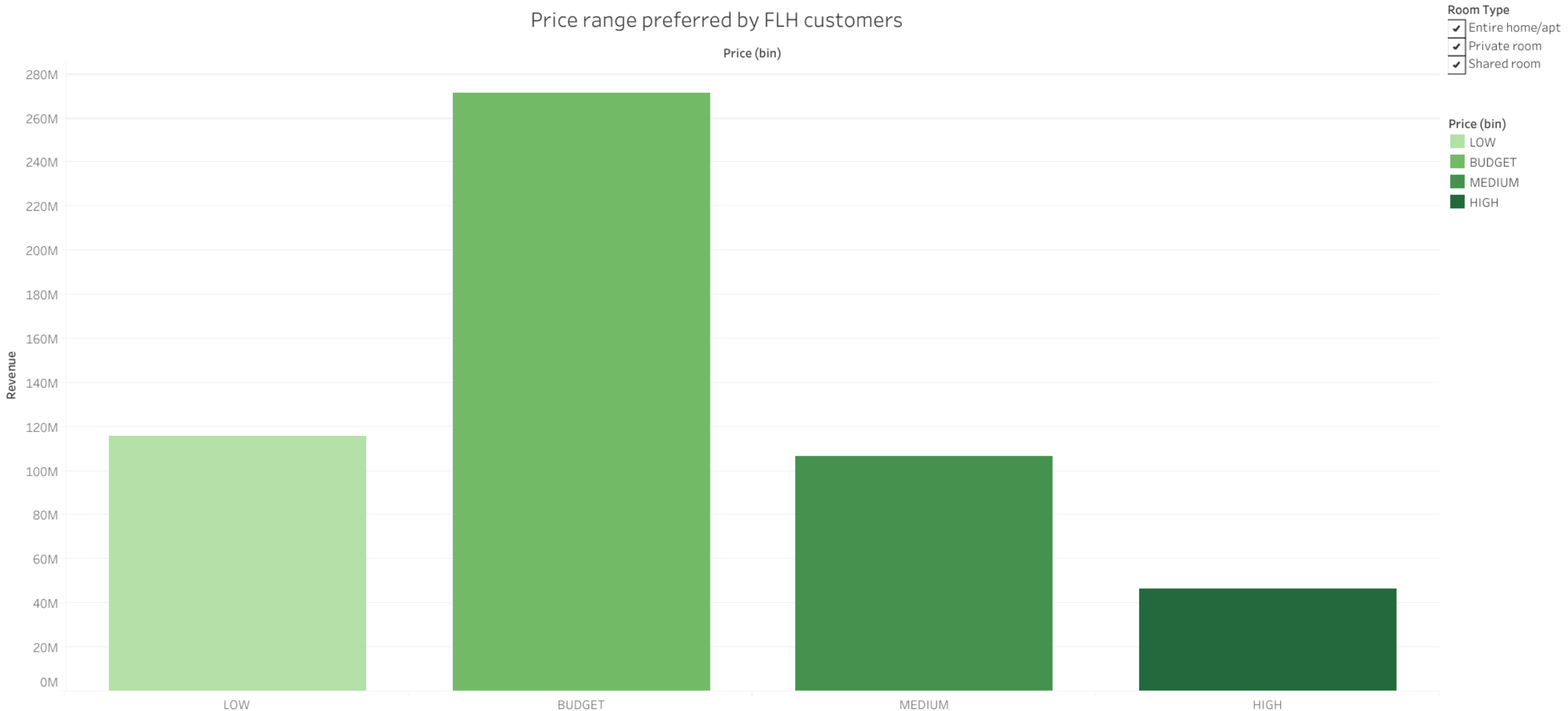
Revenue

Shared room Private room Entire home/apt

- Bronx
- Brooklyn
- Manhattan
- Queens
- Staten Island

Observation
The host listed from Brooklyn and Manhattan neighborhood group contributes 85% of the total listing, FLH needs to acquire more properties from Bronx, Queens, and Staten Island to increase the revenue. Also Entire home/ apt has more demand than a Private or Shared room so FLH needs to increase these types of properties.

Price range preferred by FLH customers

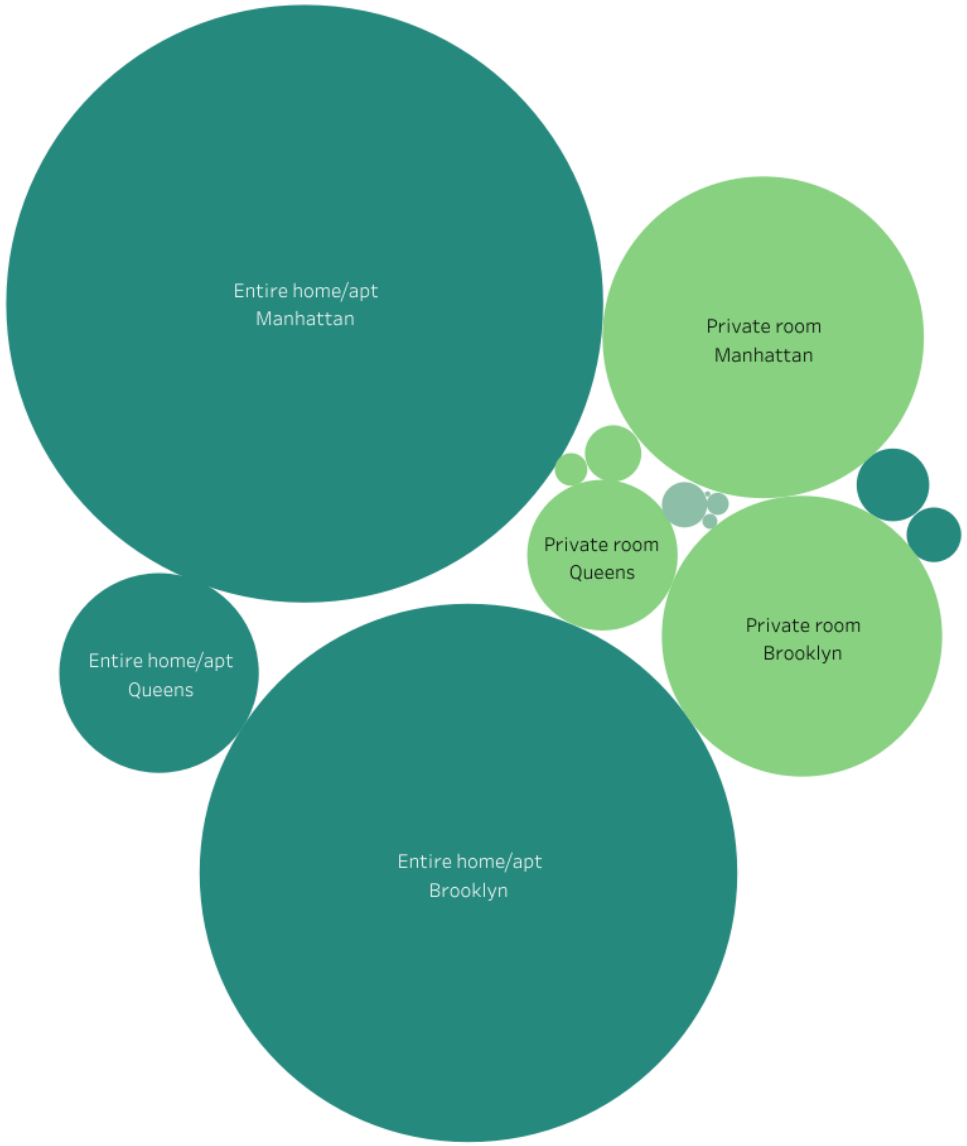


Low - (0-100); Budget - (100-200); Medium - (200-300); High - (>300$)

Observation

1) People prefer mostly budget-type rooms whose price range from 100$ -200$
2) People rarely go for the high price range
3) Need to increase more budget and medium price range rooms

Favorite Room type by Neighbourhood Group

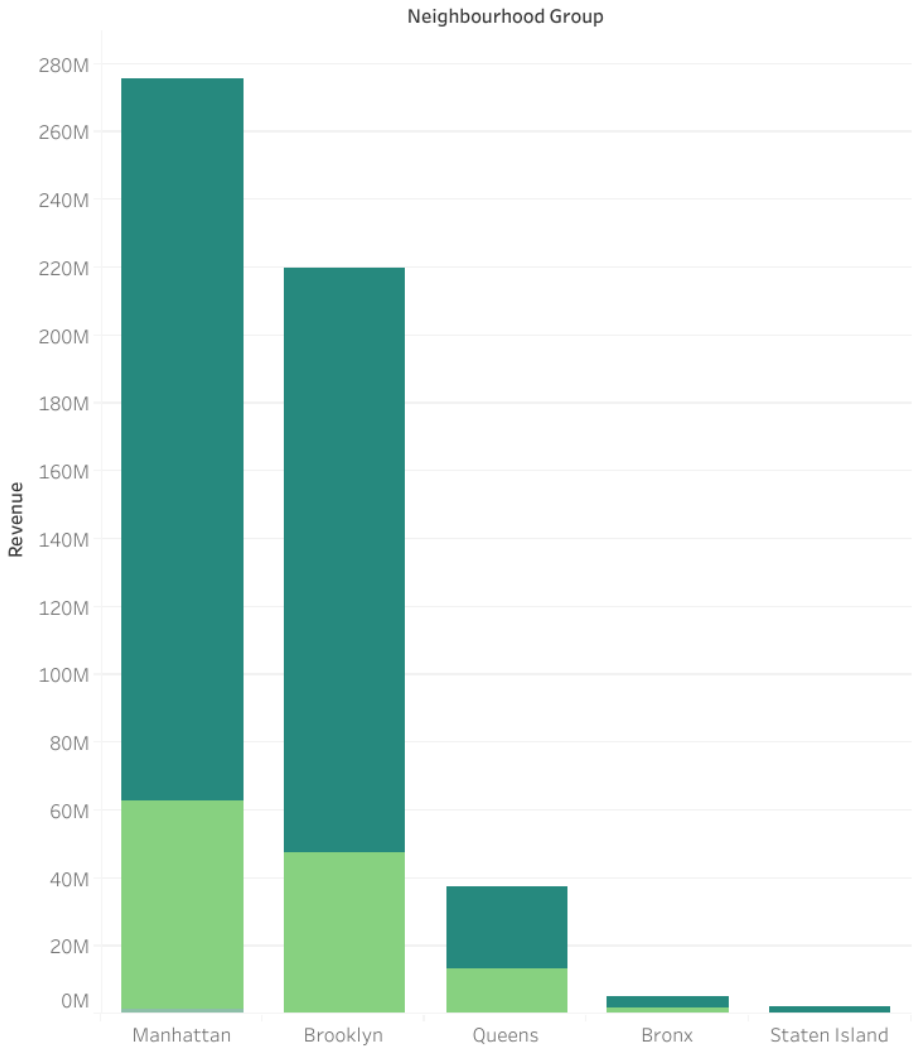


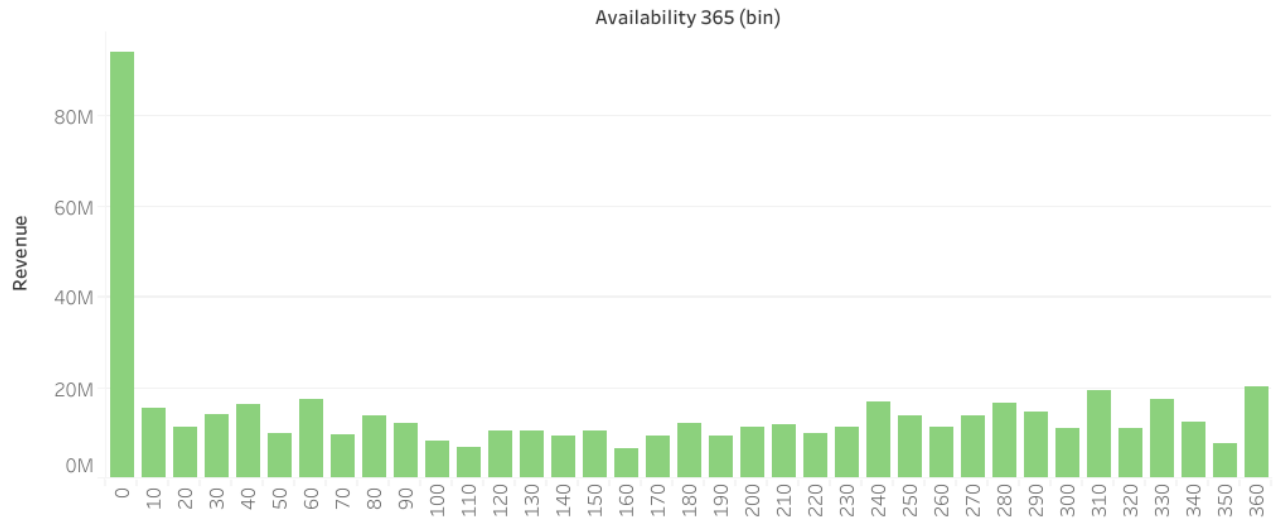
Observation:
Most Successful Type: Entire Home\Apt
Least Successful Type: Shared Room

Favorite Room type by Neighbourhood

Observation
Have to increase Entire Home\ Apt Type to generate more Revenue in low performing neighbourhood

Customers looking for most in their stays with FLH

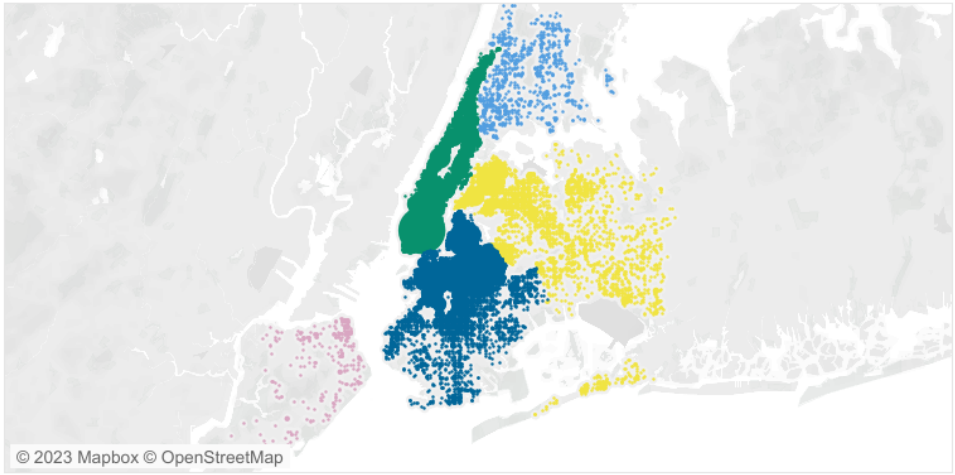


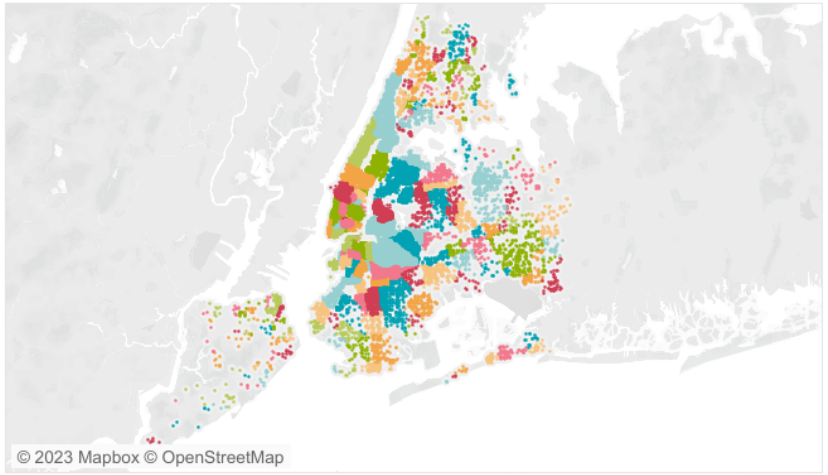


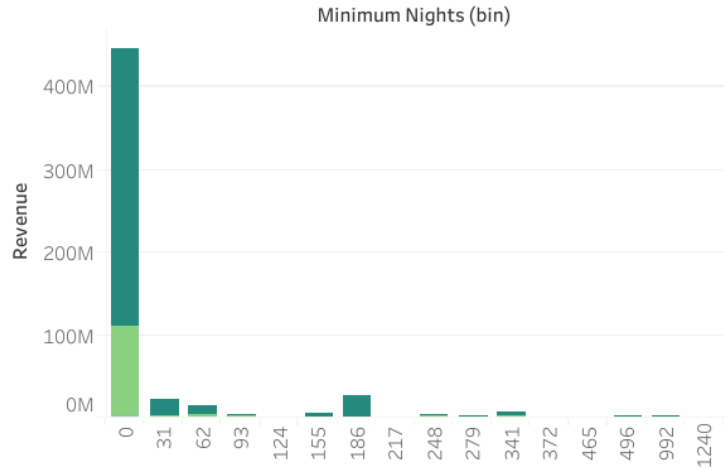
Observation:

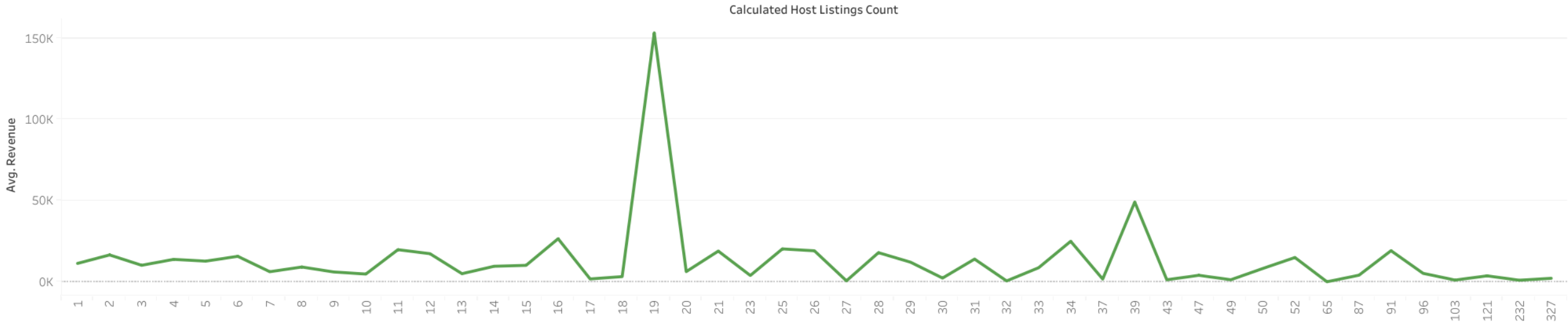
1) Most crucial thing in real estate is location. Here we can witness that the Neighbourhood group plays a crucial role.
2) About 98% of customers prefer to have a minimum night stay of fewer than 30 days.
3) Availability doesn't have that much impact on the customer preference

FLH should target to add to their service to increase revenue





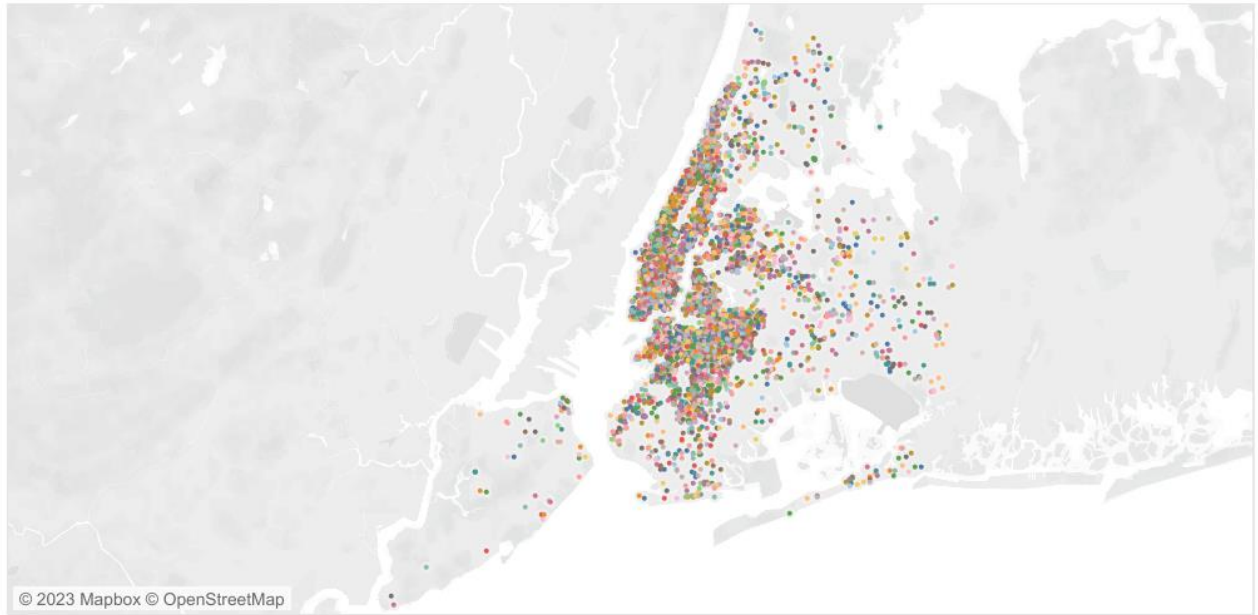




Observations:

1) The Properties listed in Manhattan(44%) and Brooklyn(40%) contributes around 85% of total New York Listings, while Queens, Bronx, and Staten Islands Contribute comparatively less, So FLH needs to target these areas and acquire more properties.
2) Staten Island and Bronx have scattered and less dense properties compared with other groups
3) About 98% of Customers opt for a minimum stay to be less than a month, so FLH needs to reduce the Minimum night stays.
4) Occasionally, listing the location on Airbnb might not be a good idea to maximize profits.

To get unpopular properties more traction



Observations:

1) "Too much of anything is good for nothing" more concentration of properties in certain neighborhood groups causes less entry into those areas

2) One of the best ways to ensure a steady stream of bookings is to make sure the FLH listing is up-to-date with proper descriptions.

3) FLH Rate changes: Even in the busiest neighborhood the price need to be adjusted with many factors like minimum night stays and availability etc...

Conclusion

- The host listed from Brooklyn and Manhattan neighbourhood group contributes 85% of the total listing, FLH needs to acquire more properties from Bronx, Queens, and Staten Island to increase the revenue. Need to increase more budget and medium price range rooms
- The Properties listed in Manhattan(44%) and Brooklyn(40%) contributes around 85% of total New York Listings, while Queens, Bronx, and Staten Islands Contribute comparatively less, So FLH needs to target these areas and acquire more properties.
- Staten Island and Bronx have scattered and less dense properties compared with other groups
- About 98% of Customers opt for a minimum stay to be less than a month, so FLH needs to reduce the Minimum night stays.