NYC Airbnb Analysis

June 26, 2022

1 The Big Apple - Analyzing Airbnb Trends in NYC

1.1 Introduction

New York City is an iconic global centerstage when it comes to art, fashion, architecture and food. It is no wonder then that it regularly features on several lists of must-visit cities and bucket lists of many travelers around the world. From couch-surfers to hostel backpackers to glam penthouses, the city has something for everyone when it comes to accommodation.

Since it's inception in 2008, Airbnb has been a game changer for tourists world over. People who previously struggled to get hotel reservations during peak tourist season and holidays suddenly had hundreds of hosts in the city open up their homes as cheaper alternatives to a hotel room. Since then, NYC has been home to well over 35,000[1] Airbnb listings in each of it's five boroughs, offering their own unique experiences.

This project takes a look at the different factors influencing Airbnb rentals in NYC. The questions going to be answered from this analyis of data include but are not limited to: - Do particular months or days of the year affect rental rates? - What are some of the priciest areas in NYC? - What types of rentals are people going for? - What is the availability of rentals in each area and how many nights are people spending on an average? - What areas of NYC get the most reviews? - What do renters advertise the most when putting up their flat for rent?

1.2 Limitations and Ethical Implications

The biggest limitation of this project is that the dataset being used belongs to 2019. Owing to the Covid pandemic in 2020, tourism and travel got disrupted and hence most data for bookings from that year onwards is heavily skewed pertaining to the different waves of Delta, Omicron, etc. For this purpose data from 2019 is being analyzed in order to get a clear picture as to what used to influence the rate of renting the most before the pandemic.

The dataset taken from InsideAirbnb.com explicitly prohibits web scraping of their data or scripts that download their data as well. They do however allow users to take the data they need in the form of csv file downloads provided that it is not republished on any major website. As such, this dataset has been made available for viewing and downloading by the website themselves. This dataset is being used purely for academic purposes and won't be published for personal gain, we can safely say there was no prohibitions to the data that is being used for this project.

Any negative outcomes from this project is unlikely to cause any large scale effect in rental rates as NYC remains an immensely popular destination for tourists.

1.3 Libraries Used

Pandas and Numpy have been used to set up and read files and sort array data. Seaborn and matplotlib have been used to code graphs and bar charts. Wordcloud has been used to generate a word-based representation of data.

[1]:
```
#setting up libraries to be used
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
from wordcloud import WordCloud

%matplotlib inline
```

1.4 Dataset Used

The following dataset contains information for all listings on Airbnb for the year 2019. This includes factors such as - prices per night, type of accommodation, number of reviews for each listing, the area of the listing,etc. In this section, we will set up the file necessary to extract data, clean the data and represent the data visually using pandas.

[2]:
```
nyc_df = pd.read_csv("nycdata.csv")
```

[3]:
```
nyc_df.head()
```

[3]:
```
     id                                              name  host_id  \
0  2539                Clean & quiet apt home by the park     2787
1  2595                             Skylit Midtown Castle     2845
2  3647               THE VILLAGE OF HARLEM…NEW YORK !     4632
3  3831                   Cozy Entire Floor of Brownstone     4869
4  5022  Entire Apt: Spacious Studio/Loft by central park     7192

     host_name neighbourhood_group neighbourhood  latitude  longitude  \
0         John            Brooklyn    Kensington  40.64749  -73.97237
1     Jennifer           Manhattan       Midtown  40.75362  -73.98377
2    Elisabeth           Manhattan        Harlem  40.80902  -73.94190
3  LisaRoxanne            Brooklyn  Clinton Hill  40.68514  -73.95976
4        Laura           Manhattan   East Harlem  40.79851  -73.94399

         room_type  price  minimum_nights  number_of_reviews last_review  \
0     Private room    149               1                  9  2018-10-19
1  Entire home/apt    225               1                 45  2019-05-21
2     Private room    150               3                  0         NaN
3  Entire home/apt     89               1                270  2019-07-05
4  Entire home/apt     80              10                  9  2018-11-19

   reviews_per_month  calculated_host_listings_count  availability_365
```

```
0            0.21                                 6          365
1            0.38                                 2          355
2             NaN                                 1          365
3            4.64                                 1          194
4            0.10                                 1            0
```

[4]:
```
nyc_df.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
```

[5]:
```
nyc_df.shape
```

[5]: (48895, 16)

[6]:
```
##room types available
nyc_df.room_type.unique()
```

[6]: array(['Private room', 'Entire home/apt', 'Shared room'], dtype=object)

[7]:
```
#areas in the city
nyc_df.neighbourhood_group.unique()
```

[7]:
```
array(['Brooklyn', 'Manhattan', 'Queens', 'Staten Island', 'Bronx'],
      dtype=object)
```

```
[8]: #number of hosts
nyc_df.host_id.nunique()
```

```
[8]: 37457
```

```
[9]: #number of listings
nyc_df.id.nunique()
```

```
[9]: 48895
```

1.5 Cleaning data

This section deals with errors in the dataset and the various ways to clean them.

```
[10]: #check for missing data/null values
def nyc_missing(nyc_df):
    agg = 0
    for col in nyc_df.columns:
        invalid_value = nyc_df[col].isnull().sum()
        agg += invalid_value
        if invalid_value != 0:
            print (f"{col} => {nyc_df[col].isnull().sum()}")
    if agg == 0:
        print("no error")
nyc_missing(nyc_df)
```

```
name => 16
host_name => 21
last_review => 10052
reviews_per_month => 10052
```

-For the purpose of this analysis, the host name column is not important and thus can be dropped from our dataset. -We can replace the NaN values in reviews_per_month. -The column last_review can be converted to datetime datatype. This way important information such as day, month and year can be extracted for the purpose of our analysis. -Finally the last_review can also be dropped from the dataset.

```
[11]: nyc_df.drop(['host_name'], axis=1, inplace=True)
```

```
[12]: nyc_df.fillna({'reviews_per_month': 0}, inplace=True)
```

```
[13]: nyc_df['last_review'] = pd.to_datetime(nyc_df['last_review'])
```

```
[14]: nyc_df['Year_last_review'] =  nyc_df['last_review'].dt.year
nyc_df['Month_last_review'] = nyc_df['last_review'].dt.month_name()
nyc_df['Day_last_review'] = nyc_df['last_review'].dt.day_name()
```

```
[15]: nyc_df.drop(columns=['last_review'],axis=1,inplace=True)
```

```
[16]: nyc_df.duplicated().sum()
```

```
[16]: 0
```

```
[17]: ##for clarity make distinction between neighborhood_group and neighborhood␣
      ↪columns by replacing it with area
      nyc_df.rename(columns={'neighbourhood_group': 'area', 'neighbourhood':␣
      ↪'neighborhood'}, inplace=True)
```

```
[18]: nyc_df.head()
```

```
[18]:     id                                              name  host_id       area  \
      0  2539                Clean & quiet apt home by the park     2787   Brooklyn
      1  2595                             Skylit Midtown Castle     2845  Manhattan
      2  3647               THE VILLAGE OF HARLEM…NEW YORK !     4632  Manhattan
      3  3831                   Cozy Entire Floor of Brownstone     4869   Brooklyn
      4  5022  Entire Apt: Spacious Studio/Loft by central park     7192  Manhattan

        neighborhood  latitude  longitude        room_type  price  minimum_nights  \
      0   Kensington  40.64749  -73.97237     Private room    149               1
      1      Midtown  40.75362  -73.98377  Entire home/apt    225               1
      2       Harlem  40.80902  -73.94190     Private room    150               3
      3 Clinton Hill  40.68514  -73.95976  Entire home/apt     89               1
      4  East Harlem  40.79851  -73.94399  Entire home/apt     80              10

        number_of_reviews  reviews_per_month  calculated_host_listings_count  \
      0                 9               0.21                               6
      1                45               0.38                               2
      2                 0               0.00                               1
      3               270               4.64                               1
      4                 9               0.10                               1

        availability_365  Year_last_review Month_last_review Day_last_review
      0              365            2018.0           October          Friday
      1              355            2019.0               May         Tuesday
      2              365               NaN               NaN             NaN
      3              194            2019.0              July          Friday
      4                0            2018.0          November          Monday
```

1.6 Checking for outliers in the dataset

Outliers in the dataset can include bookings that are more than the allowed period of time, price ranges that are too low or high, availability of flats that have already been booked, etc. For the purpose of getting accurate results from our analysis, we will have to search for and deal with outliers first.

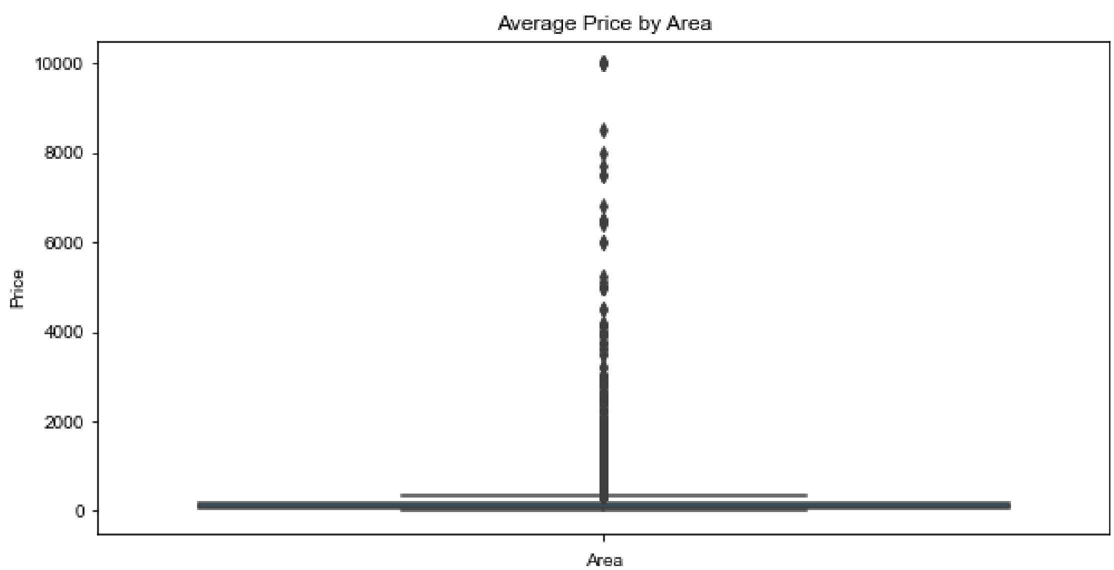
[19]:
```
nyc_df.agg(
    {'price': ['mean', 'median', 'min', 'max', 'count']})
```

[19]:

	price
mean	152.720687
median	106.000000
min	0.000000
max	10000.000000
count	48895.000000

From the output we can see that the min value is set to 0 which should not be possible as Airbnb does not host free rooms. We can plot the data as a boxplot to better understand where outliers lie.

[20]:
```
plt.figure(figsize=(10,5))
ax = sns.boxplot(y='price', data=nyc_df).set_title('Average Price by Area')
sns.set_theme(style='white')
plt.xlabel('Area')
plt.ylabel('Price')
plt.show()
```



As expected, there are several rooms listed for 0 and even some for 10,000 which seem absurd as standard price for a night. One of the ways we can clean this data and remove outliers is by implementing a floor and a ceiling to prevent skewing of data too wildly. Data can be extracted between the percentages of 10%-90%, which will allow us to drop extremely low and high values in the dataset.

[21]:
```
nyc_floor = nyc_df['price'].quantile(0.10)
nyc_floor
```

[21]: 49.0

[22]:
```python
nyc_ceiling = nyc_df['price'].quantile(0.90)
nyc_ceiling
```

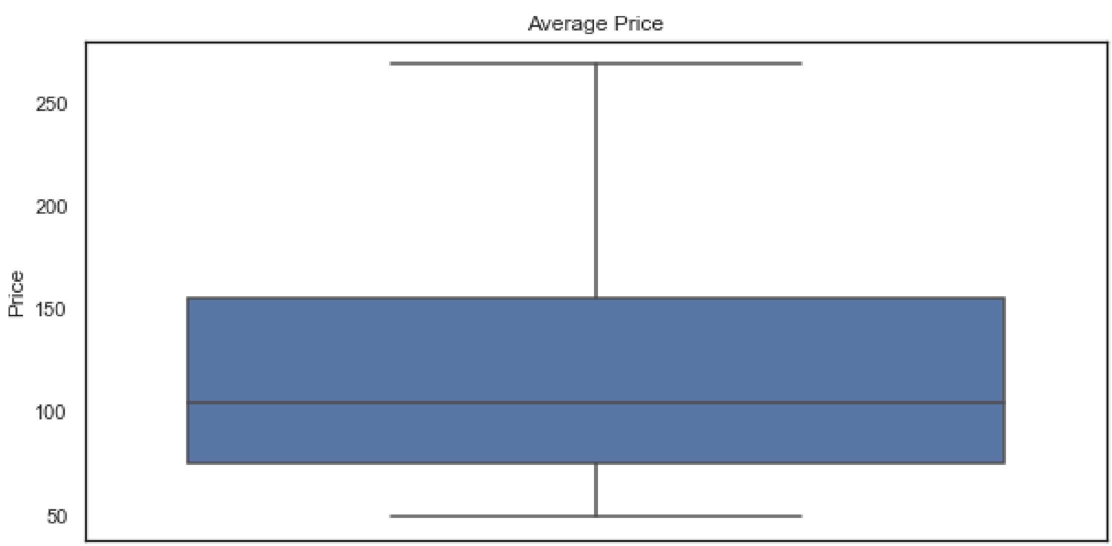
[22]: 269.0

The floor of the prices is 49 and the ceiling is 269, which seem much more reasonable prices for a room. We can use this to filter out any results less than 49 and higher than 269 a night.

[23]:
```python
nyc_df = nyc_df.drop(nyc_df[nyc_df.price < nyc_floor].index)
```

[24]:
```python
nyc_df = nyc_df.drop(nyc_df[nyc_df.price > nyc_ceiling].index)
```

[25]:
```python
#create new plot to check if implementation successful
plt.figure(figsize=(10,5))
ax = sns.boxplot(y='price', data=nyc_df).set_title('Average Price')
sns.set_theme(style='white')
plt.ylabel('Price')
plt.show()
```

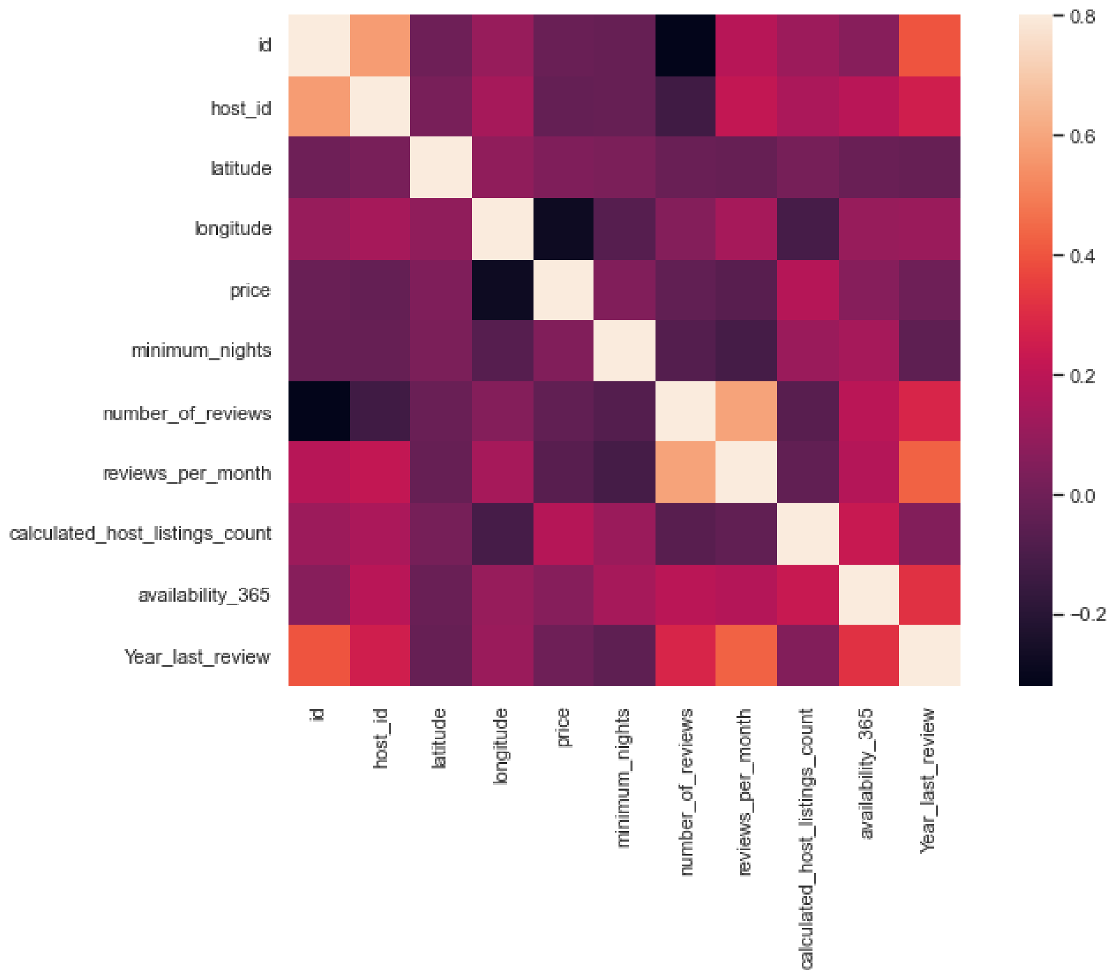


Our data is now clean and can be used for analysis.

1.7 Data Analysis and Representation

This section deals with the analysis of the cleaned data and it's representation in the form of charts, graphs and maps. The first step in our analysis is to find any strong correlation between the columns of the dataset to see whether they impact each other in a significant way. A seaborn heatmap is implemented for this purpose.

[26]:
```
nyc_corr = nyc_df.corr()
f, ax = plt.subplots(figsize=(12, 7))
sns.heatmap(nyc_corr, vmax=.8, square=True);
```



It can be observed from the heatmap that most of the correlation coefficients are lower than 0.5, which is not significant to our analysis. The only correlation of note is between minimum_nights and calculated_host_listings_count which can be compared in analysis. Thus we can keep most of our data analysis focused on independant columns themselves as most of them seem to have no significant effect on each other.

1.7.1 Factor 1 - Total Listings by Each Area

This section will give us a better idea of the total number of listings per area in NYC. The density of listings in an area is crucial to whether or not it performs better in rate of bookings and reviews. We can use a pivot table to get the statistics.

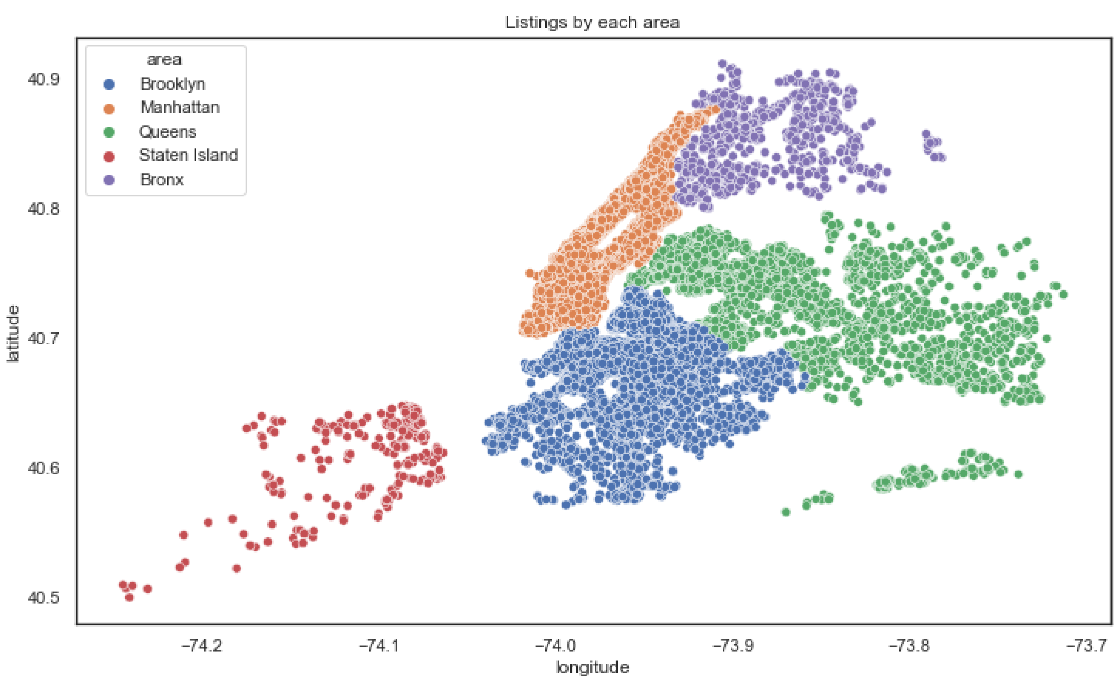
[27]:
```python
nyc_pivot = pd.pivot_table(nyc_df, index=['area'], values='id',␣
 ↪aggfunc=['count'],
                              margins=True, margins_name='Total Listings')
nyc_pivot
```

[27]:

	count
	id
area	
Bronx	758
Brooklyn	16373
Manhattan	17513
Queens	4413
Staten Island	278
Total Listings	39335

The data from the pivot table shows us that Manhattan has the most listings of all 5 areas, followed closely by Brooklyn. The other three areas seem nowhere near in terms of density of listings. A crude hypothesis from this can be that Manhattan might be the leading area in performance for prices, minimum nights and reviews. This pivot list can be better visualized using a scatterplot.

[28]:
```python
#total listing by each area
plt.figure(figsize = (12,7))
sns.scatterplot(x="longitude" ,y="latitude", data=nyc_df, hue = "area", alpha =␣
 ↪1)
plt.title("Listings by each area")
plt.show()
```

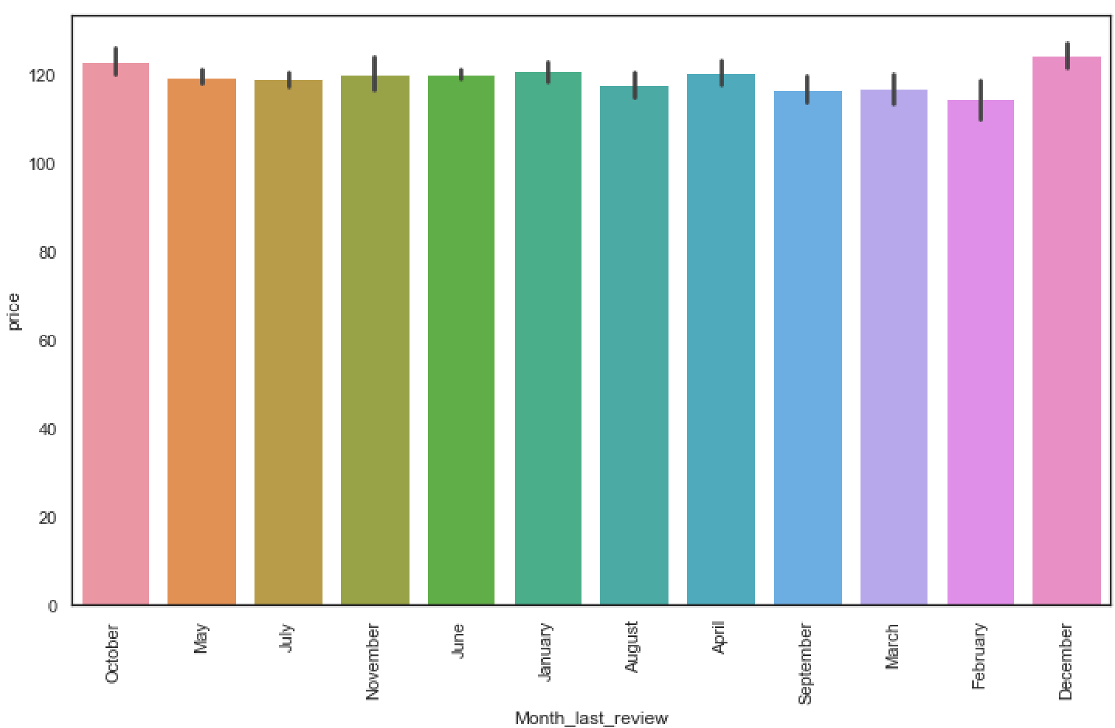


The scatterplot confirms the pivot table findings, with Manhattan and Brooklyn being the most densely populated with listings. Staten island can be clearly seen to have the least amount of listings. A quick search[2] on the most famous touristy places in NYC gives us a picture of why this might be: - The highest rated tourist attractions of the Empire State Building, the Statue of Liberty, Central Park and Grand Central Station are all located in Manhattan. - Brooklyn also has it's fair share of tourist traffic given that it has the Brooklyn Bridge, Coney Island and Barclays Center. - Both of the largest airpost in NYC are located in these two areas. Manhattan has La Guardia(LGA) while Brooklyn has JFK International. This would be favorable to tourists on a short stay who wouldn't want to live far away from city center.

1.8 Factor 2 - Influence of Particular Months and Days

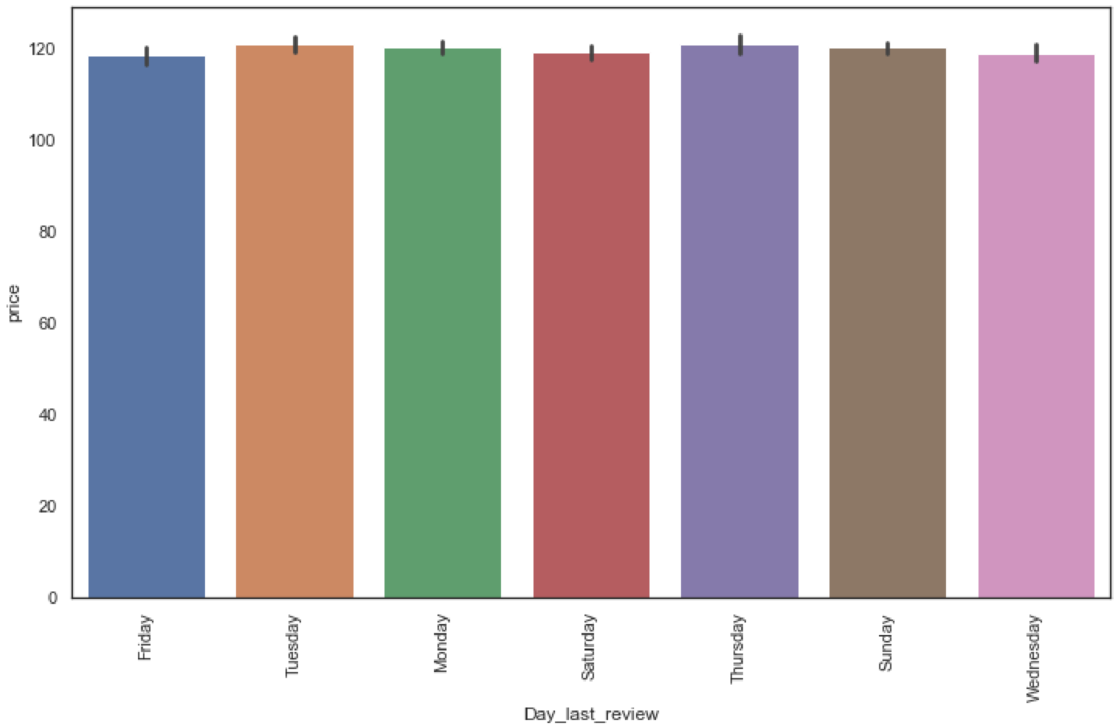
This section deals with analyzing whether Airbnb sees more bookings during a particular time of the year and if so, do travelers prefer a certain day of the week to book places. The initial hypothesis for this factor is Airbnb would see a spike during major American holidays[3] such as Thanksgiving and Christmas. We can also assume people would tend to book more during weekends as they would have a break from work/school.

[29]:
```
#months when prices are costliest
plt.figure(figsize = (12,7))
sns.barplot(x=nyc_df['Month_last_review'],y=nyc_df['price'])
plt.xticks(rotation='vertical')
plt.show()
```



As expected, December has a higher spike than the rest of the months owing to an increase in prices, which means an increase in bookings. However the rest of the months don't seem to have any significant deviation from each other.

[30]:
```
#days when prices are costliest
plt.figure(figsize = (12,7))
sns.barplot(x=nyc_df['Day_last_review'],y=nyc_df['price'])
plt.xticks(rotation='vertical')
plt.show()
```



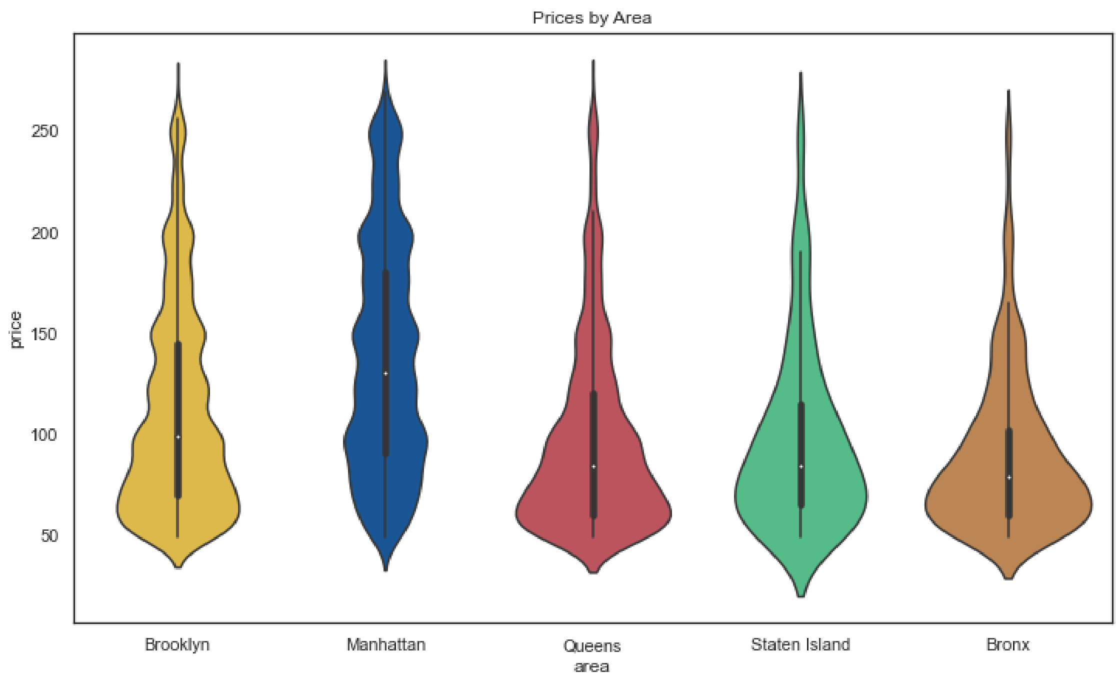
Our initial hypothesis of bookings being more popular during the weekends is refuted as there is not major difference in the prices of bookings over all days of the week.

1.9 Factor 3 - Priciest Areas to Book an Airbnb

This section deals with analyzing whether a particular area of NYC is pricier than the others. As seen in our analysis above, Manhattan has already proven to be a pretty popular place to rent. The initial hypothesis from this can be that Manhattan must be the priciest area in NYC. We can use a violin plot to visualize this data.

[31]:
```
#price distribution by each area
nyc_col = {'Brooklyn': '#f5c431', 'Manhattan': '#0555ab', 'Queens': '#cc434f',␣
↪'Staten Island': '#43cc8a', 'Bronx': '#cc8543'}
```

[32]:
```
plt.figure(figsize=(12,7))
ax = sns.violinplot(x="area", y="price", data=nyc_df, palette=nyc_col).
 ↪set_title('Prices by Area')
plt.show()
```



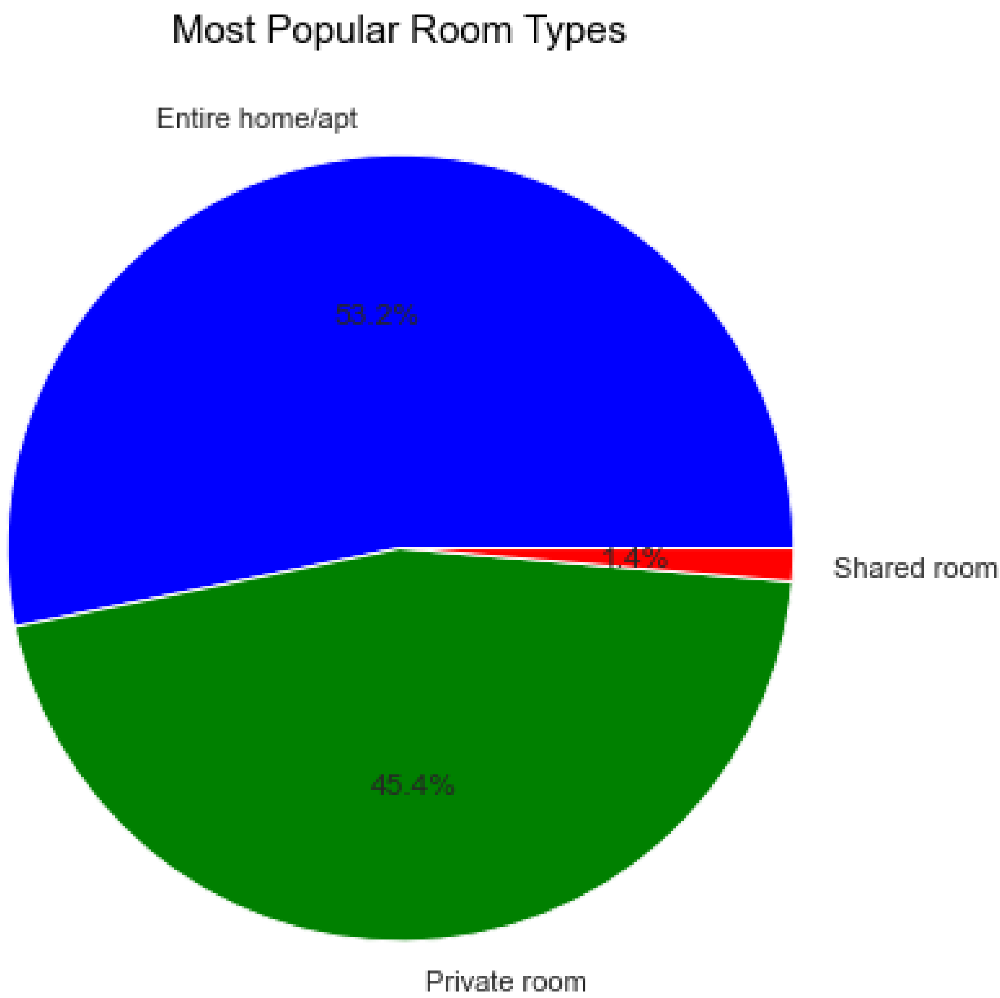
As expected our initial hypothesis was correct, with Manhattan having the highest range of prices for bookings throughout. Comparatively, the rest of the areas have more or less an equal curve, with Bronx looking to be the most affordable area.

1.10 Factor 4 - Type of Rentals

This section deals with analyzing the type of rentals that seem to be booked the most on Airbnb. The app itself offers a variety of rental units ranging from single rooms to cabins to even boats. Since NYC attracts such a wide variety of tourists, our initial hypothesis is that a single private room would be the most popular choice for bookings. Given that the price of a single room is much less than that of an entire apartment or house, travelers might pick that as a first choice to stay in their budget while in the city.

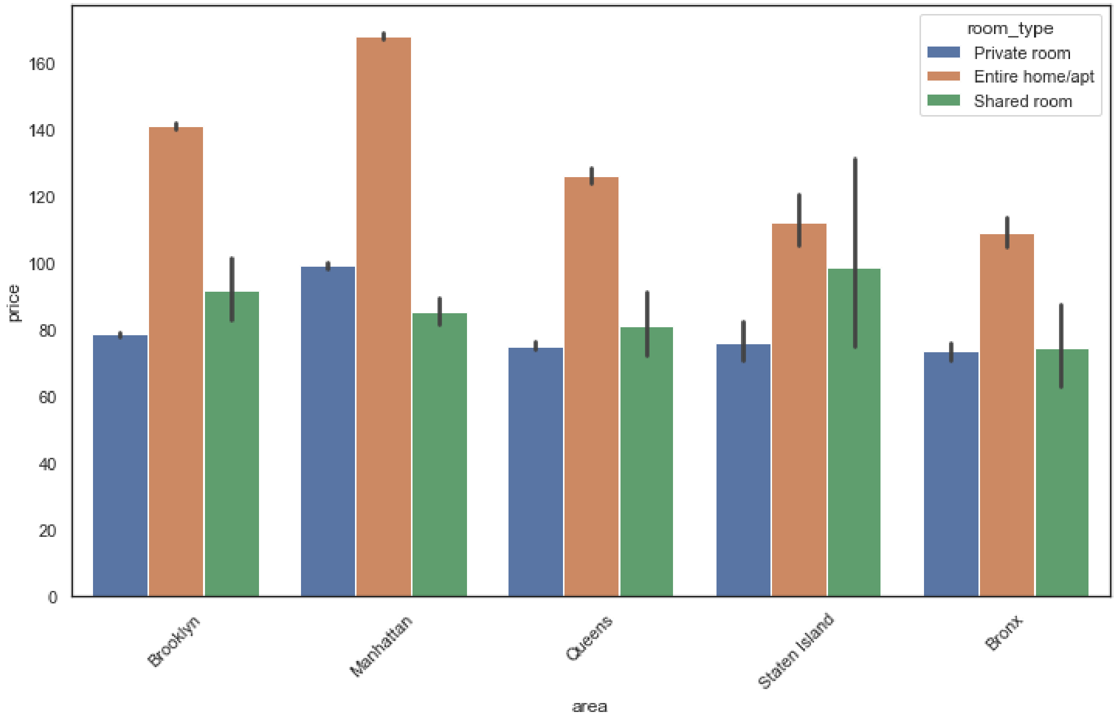
[33]:
```
#room types most popular
desc = nyc_df.room_type.value_counts().index
col = ['blue','green','red']
explode = [0,0,0]
sizes = nyc_df.room_type.value_counts().values

plt.figure(0,figsize = (12,7))
plt.pie(sizes, explode=explode, labels=desc, colors=col, autopct='%1.1f%%')
```

```
plt.title('Most Popular Room Types',color = 'black',fontsize = 15)
plt.show()
```



Even though our initial hypothesis is refuted, it is done so by a small margin of 7% as the preference for an entire house and a private room in a house is nearly equally split. It can be assumed that large families might prefer entire apartments/houses while couples or single people traveling prefer the private room. Shared rooms seem to have the smallest amount of bookings as most people might prefer paying a bit extra for privacy and security.

[34]:
```
#room type preferred by each area
plt.figure(figsize=(12,7))
sns.barplot(x = "area", y = "price", hue = "room_type", data = nyc_df)
plt.xticks(rotation=45)
plt.show()
```



Consistent with our analysis of the most popular room type throughout NYC, each separate areas themselves also seem to prefer bookings for entire apartments/houses. Manhattan and Brooklyn in particular have a large margin of tourists preferring to book entire apartments/houses.

1.11 Factor 5 - Room Availability and Minimum Nights Spent Per Listing

This section deals with analyzing the availaibility of rooms throughout the year by each area and also the minimum number of nights guests spent at their bookings. Our initial hypothesis for the availability of rooms is that Manhattan and Brooklyn might have low rates of availability seeing as they are the most popular destinations for bookings. From our analysis above, we already know that Bronx and Staten Island are among the cheaper areas to stay. Therefore we can base our initial hypothesis for minimum nights on this by saying that people who book listings in those two areas would tend to spend more nights on average than the people who book listings in Manhattan as listings there would be pricier.

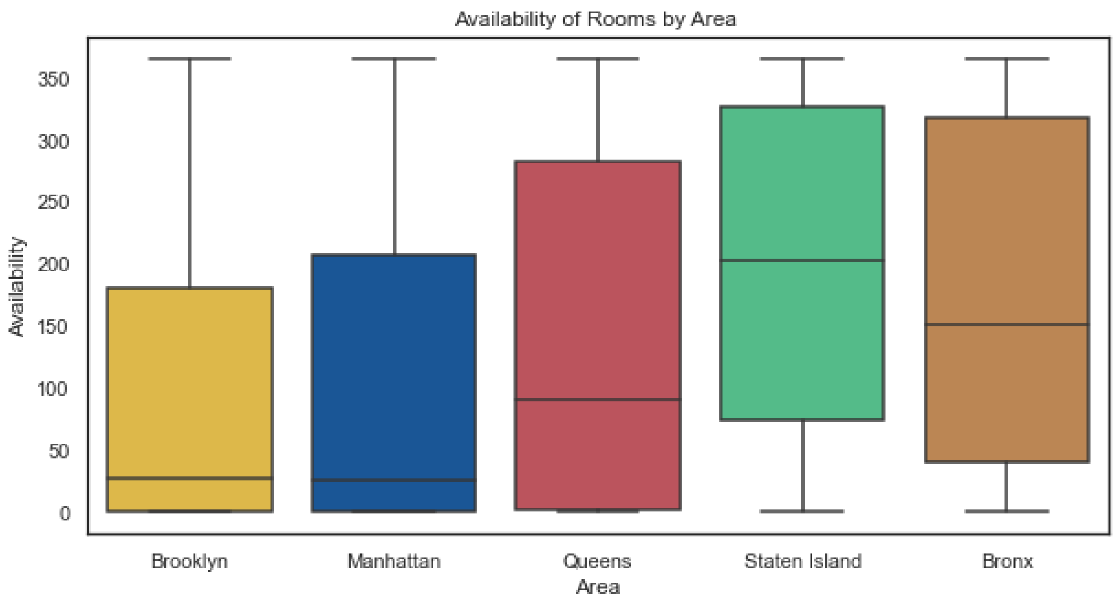
[35]:
```
#room availability by each area
nyc_df.groupby(['area'])['availability_365'].mean()
```

[35]:
```
area
Bronx            169.184697
Brooklyn          97.789837
Manhattan        103.723063
Queens           141.524813
Staten Island    196.471223
```

```
Name: availability_365, dtype: float64
```

As expected, the number of listings available in Staten Island and Bronx are greater than those in Manhattan. Availaiblity in this case would mean more vacant rooms/houses in that specific area. This shows us that even though Manhattan and Brooklyn are the most densely populated with the most number of listings available, the number of listings vacant/waiting to be booked are the highest in Staten Island and Bronx. This lines up with our earlier analysis of Manhattan being the most sought out place to book an Airbnb.

[36]:
```
plt.figure(figsize=(10,5))
ax = sns.boxplot(data=nyc_df, x='area', y='availability_365', palette=nyc_col).
 ↪set_title('Availability of Rooms by Area ')
sns.set_theme(style='white')
plt.xlabel('Area')
plt.ylabel('Availability')
plt.show()
```

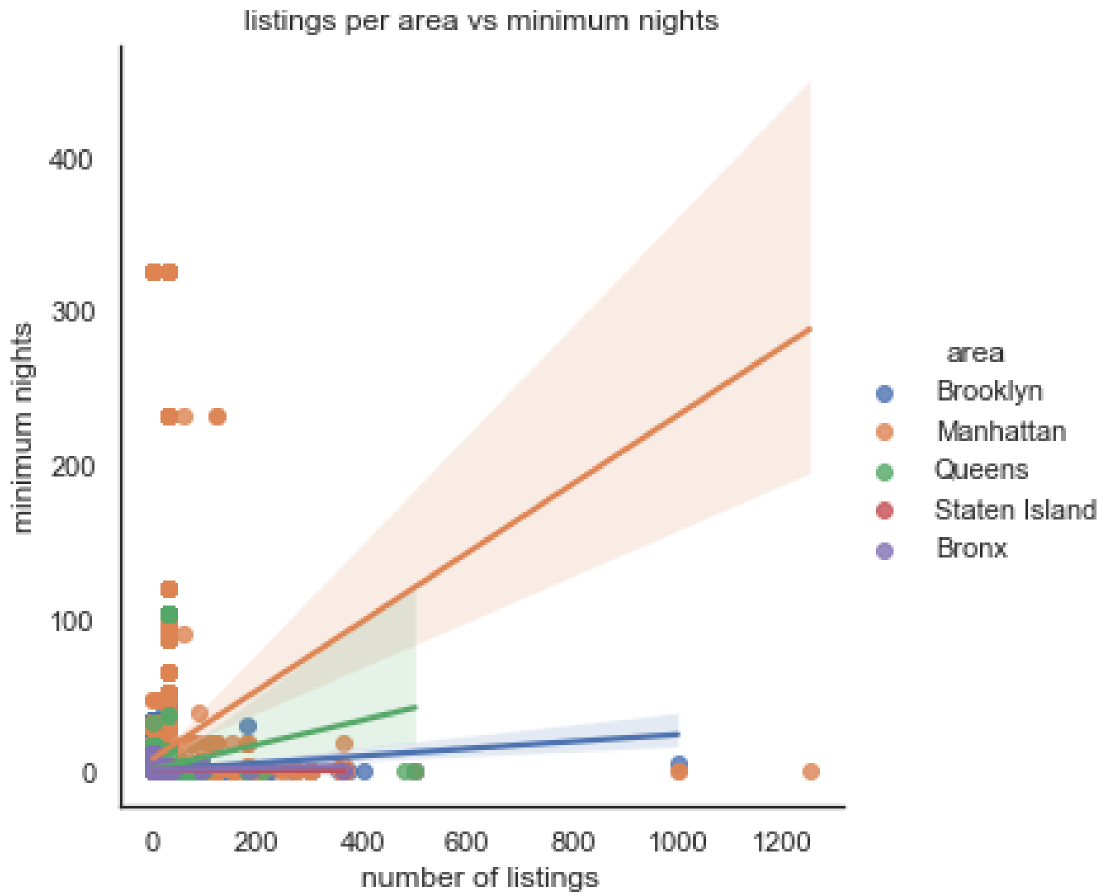


Data visualized through boxplots show us the average number of rooms vacant in both Manhattan and Brooklyn are low, whereas those in Staten Island particularly are high.

[37]:
```
#min number of nights in each area
plt.figure(figsize=(18,18))
sns.
 ↪lmplot(x='minimum_nights',y='calculated_host_listings_count',hue="area",data=nyc_df)
plt.xlabel('number of listings')
plt.ylabel('minimum nights')
plt.title('listings per area vs minimum nights')
```

```
plt.show()
```

<Figure size 1296x1296 with 0 Axes>



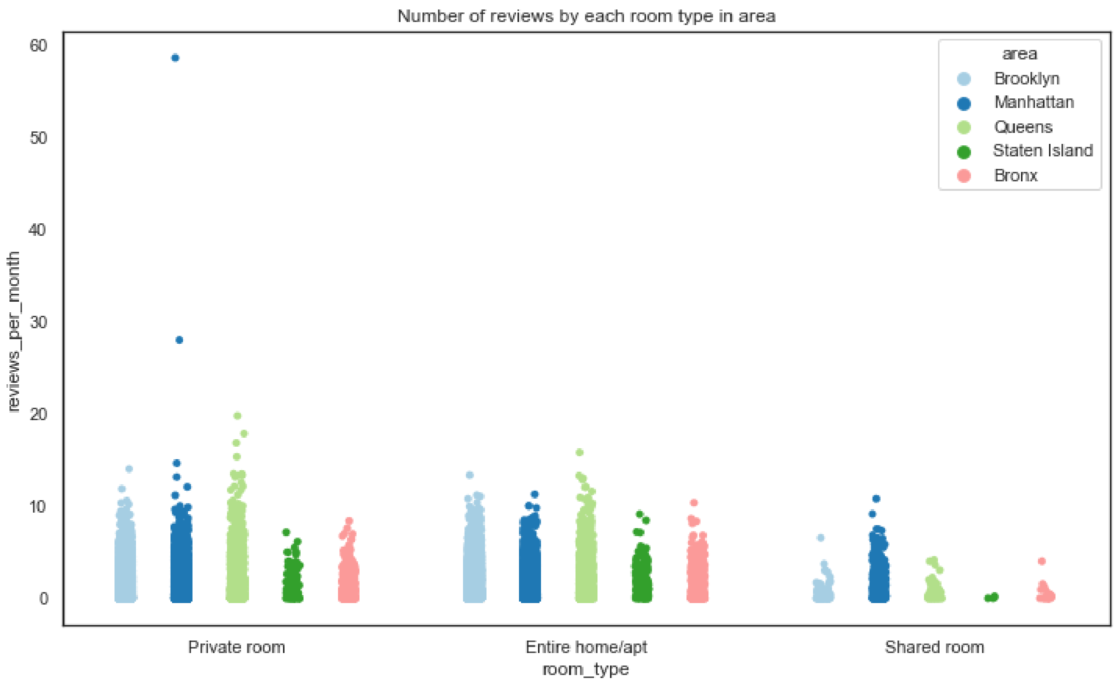
Our initial hypothesis of Staten Island and Bronx having a higher average of guests staying more nights is refuted as Manhattan seems to dominate this category as well. It can be assumed that though prices are indeed higher for Manhattan, being at the center of major tourist attractions as well as transport seems to be a priority for most guests.

1.12 Factor 6 - Guest Opinions and Advertising of Listings

This section deals with the analysis of the number of reviews left by guests for their bookings by each area. Our initial hypothesis is since Staten Island and Bronx have the most number of vacant listings, they have the least number of reviews available. Since a general analysis will be too broad to get a clear picture, we can break down the number of reviews by each room type left by the guests. We will also take a look at how hosts on Airbnb advertise their listings by implementing a wordcloud to figure out the most popular terms used.

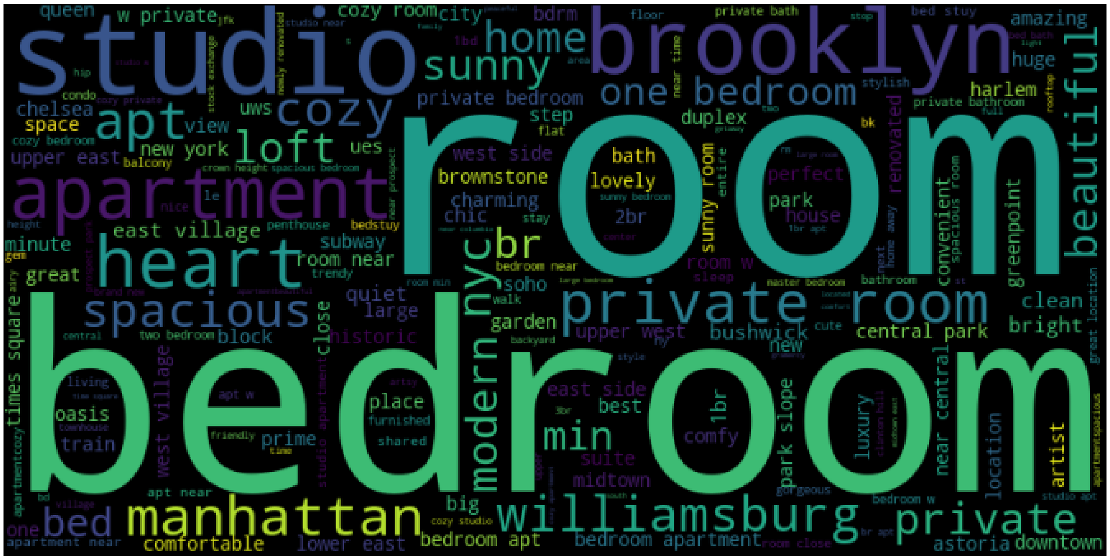
[38]:
```python
#no of reviews by each room type in areas
f,ax = plt.subplots(figsize=(12,7))
ax= sns.
 ↪stripplot(x='room_type',y='reviews_per_month',hue='area',dodge=True,data=nyc_df,palette='Pa
ax.set_title('Number of reviews by each room type in area')
```

[38]: Text(0.5, 1.0, 'Number of reviews by each room type in area')



Our initial hypothesis of Staten Island and Bronx having the least number of reviews stands correct as the listings in those areas are most likely to be vacant compared to other areas. Less guests mean less reviews. From the strip plot above, we can see that reviews are once again split equally between private rooms and entire homes/apartments. This ties in with our earlier analysis of a very small margin between the two types of bookings available.

[39]:
```python
#most popular advertised words in listings
nyc_lst = nyc_df['name'].tolist()
text = ""
for item in nyc_lst:
    text += str(item).lower()
wordcloud = WordCloud(width = 800, height = 400).generate(text)
plt.figure(figsize = (12,7))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



A wordcloud analysis on the ‘name’ section of the listings gives us the most popular terms being used by hosts to advertise their listings. From the figure above, we can see that the type of accommodation being listed is a priority while advertising, with room, bedroom and studio being the biggest words. Brooklyn seems to be the most heavily advertised area in these listings. A special mention can also be given to the term ‘heart’ which ties in to our earlier analysis of people spending an average of more nights in Manhattan, which seems to be at the heart of all the tourist locations. Advertising this would certainly drive up booking rates.

1.13 Conclusions

At the beginning of the project, we had 6 factors that we decided to take a look at while analyzing trends in Airbnb bookings in NYC. After a thorough analysis on each of those factors, we can now answer the questions we had pertaining to them.

1.13.1 - Do particular months or days of the year affect rental rates?

Our initial hypothesis for this question was that holidays and weekends would be the most popular times to book a listing on Airbnb. Our analysis supports the hypothesis for months, as December seems to be the most booked and also the priciest month of the year. However our hypothesis of weekends having an edge over other days of the week seemed to be wrong as there is no significant difference in the price of bookings throughout any days.

1.13.2 - What are some of the priciest areas in NYC?

For this question, we first had to tackle the data for what area in NYC had the most number of listings. Doing so gave us an idea of what area was the most popular and therefore would be much pricier. Our initial analysis showed us that by far Manhattan and Brooklyn had the most number of listings by area. Further analysis on data by pricing backed up our hypothesis that Manhattan was indeed also the pricier of all the 5 areas.

1.13.3 - What types of rentals are people going for?

Airbnb has a variety of rentals available on their app and our initial hypothesis for this was private rooms would be the most popular type of rentals. This hypothesis was based on the fact that private rooms most resemble a hotel room which satisfies a guest's need for short stays. Our analysis showed a surprising almost even split between private rooms and entire houses/apartments, with the latter having a slight edge, indicating people were prepared to splurge a little more to have the privacy of an entire house to themselves.

1.13.4 - What is the availability of rentals in each area and how many nights are people spending on an average?

Going by our analysis so far in the project, we had a clear picture that Manhattan was the most popular destination in NYC for bookings. Therefore we would now have to look at the areas in NYC where the density of listings was the least to hypothesize that those areas would have the most number of listings available to book. This hypothesis turned out to be correct as both Staten Island and Bronx were the leading areas with most vacant bookings and they were also previously found to be the least densely populated with listings. For the second part of the question, we assumed that since the prices of the listings in Staten Island and Bronx was the lowest, it would persuade people to stay more nights on average than compared to the other pricier areas. This hypothesis turned out to be wrong as once again, Manhattan seemed to be the choice for guests to spend the most number of nights at. We attributed this to the location being central to all the well known tourist attractions.

1.13.5 - What areas of NYC get the most reviews?

Basing our hypothesis on the analysis of the availability of rentals, we assumed that Staten Island and Bronx would have the least amount of reviews. This was based on the fact that fewer people booking listings in these areas would also lead to fewer reviews. This hypothesis was indeed correct. Queens and Manhattan had the highest number of guests leaving reviews than other areas. We also decided to analyse this data across all room types and found the data split consistent between private rooms and entire apartments/houses. Shared rooms seemed to have the least reviews indicating them being the least preferred type of accommodation, verifying our analysis from earlier.

1.13.6 - What do renters advertise the most when putting up their flat for rent?

The 'name' column in the dataset gives us the description of the way listings were primarily advertised to the guests on the app. Doing a wordcloud analysis on this column gave us a picture of the most frequently used terms that hosts used. Out of these, the biggest ones understandably were 'ROOM', 'BEDROOM' and 'STUDIO' which are used to describe the type of accommodation that was being listed. As guests would prioritize this information above all others, it makes sense they were the most commonly used terms. Amongst areas, 'BROOKLYN' appeared to be the most advertised place in the listings, indicating the density of listings from that area. Other factors that seemed to be of importance were terms like 'HEART' indicating the listing was central to popular tourist avenues and 'COZY', 'SPACIOUS', 'BEAUTIFUL' and 'MODERN' which would catch user attention to the specialities of the listings.

Overall, this dataset has helped us analyze how the Airbnb rental market functions in a large city like NYC. Hosts looking to advertise their listing on Airbnb can use this analysis to increase or decrease their prices depending on the various factors affecting each area of NYC. Conversely, guests looking to book a listing on Airbnb can look at this data while deciding what would be the best bang for their buck in terms of space, price and ease of access to popular spots around NYC.

1.14 References

1. Statista. 2022. New York: monthly number of Airbnb listings 2020 | Statista. [online] Available at: https://www.statista.com/statistics/1265037/new-york-monthly-airbnb-listings/#:~:text=Number%20of%20Airbnb%20listings%20in%20New%20York%202010%2D2020%2C%20by%20month&text=As%20of%20October%202020%2C%20there,Airbnb%20listings%20in%20the%20city. [Accessed 26 June 2022].

2. Guides, T. and York, N., 2022. 27 Top Tourist Attractions in New York City. [online] Touropia. Available at: https://www.touropia.com/tourist-attractions-in-new-york-city/ [Accessed 26 June 2022].

3. Lavie, J. and Lavie, J., 2022. How to get your Airbnb ready for the holiday season. [online] GuestReady's Airbnb Hosting Blog. Available at: https://www.guestready.com/blog/airbnb-holiday-season/ [Accessed 26 June 2022].

4. Insideairbnb.com. 2022. Adding Data To The Debate. [online] Available at: http://insideairbnb.com/get-the-data/ [Accessed 26 June 2022].