# Data set

|  |  |  |  |
| --- | --- | --- | --- |
| **File** | **Dimension** | **Description** | **Columns** |
| **listing.csv** | (77096, 28) | Consists of many columns to describe details of listings | id 77063  name 74666  summary 69094  longitude 77063  latitude 77063  space 48680  description 72878  instant\_bookable 2  neighborhood\_overview 38737  neighbourhood\_cleansed 33  host\_id 50091  host\_name 13651  host\_since 3123  host\_response\_time 5  street 1396  zipcode 31965  review\_scores\_rating 56  property\_type 42  room\_type 3  accommodates 17  bathrooms 23  bedrooms 15  beds 20  reviews\_per\_month 950  amenities 68854  cancellation\_policy 6  number\_of\_reviews 369  price 847 |
| **calendar.csv** | (28139675, 4) | Constists of details of listings and itss availability and its price | listing\_id 77095  date 366  available 2  price 2082 |
| **reviews.csv** | (1176072, 6) | consits of reveiws for each listing | listing\_id 57553  id 1176028  date 2958  reviewer\_id 979035  reviewer\_name 116785  comments 1131802 |

# Prelimitary Analysis/ Data Exploration

# General questions

1. How many Airbnb listings
2. The average listing
3. Entire place breakdown
4. Number of bedrooms
5. Studio aparments
6. Best locations
7. Top 30 studios

# Customer Analysis

1. What is the busiest time to visit a city
2. What is the busiest neighborhood
3. Seasonal pattern of airbnb housing price
4. Which borough has a better price performance ratio (being able to enjoy great experience with limited budgets?)
5. When is thee best time to place your order in high reason?
6. Analyzing trust factor
7. Rating host by data analysis
8. Determine what kinds of info host provide in their profiles to reduce uncertainty and signal trustworthies
9. Online reputation

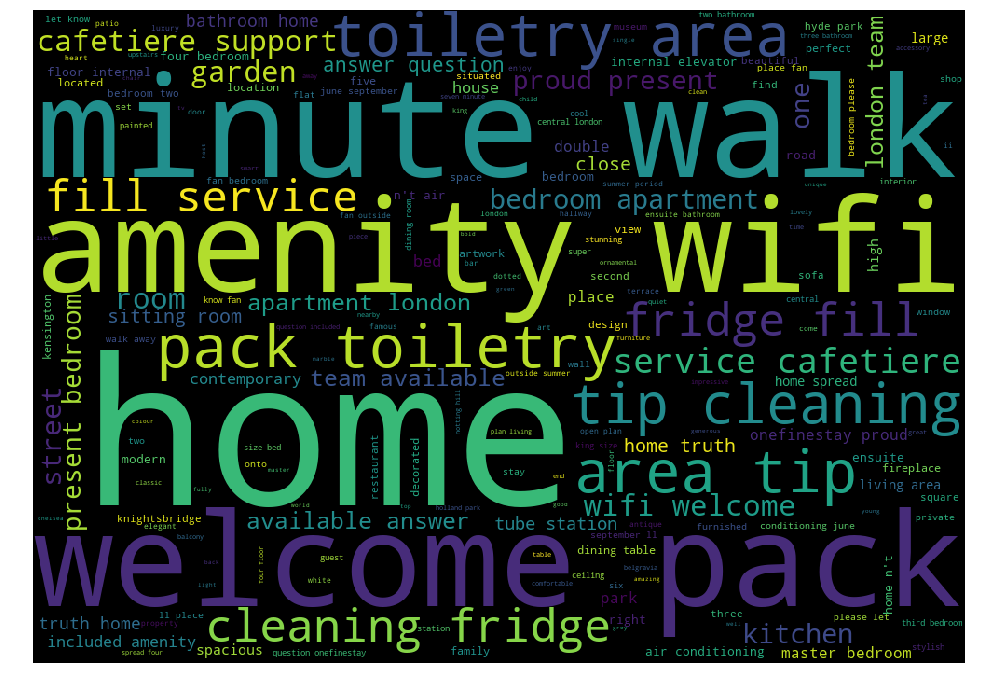
# Host Analysis

Question:

1. Predict airbnb price
2. What causes difference in price of listing

## Summary

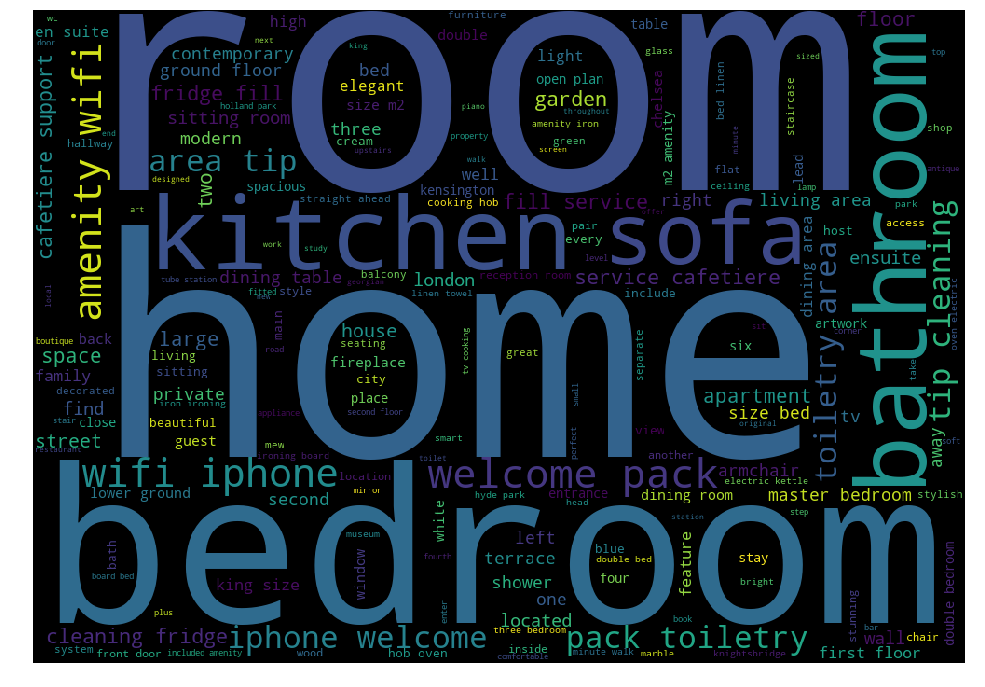
1. Now that we’ve seen how property types and room types along with neighborhood can effect the listing prices, let’s investigate the summary. Top 100 listings chosen on the basic of price to find out what all common words are used by hosts while posting a listing on airbnb. The wordcloud is employed to visualize the most common words these hosts utilized to describe their listings on airbnb



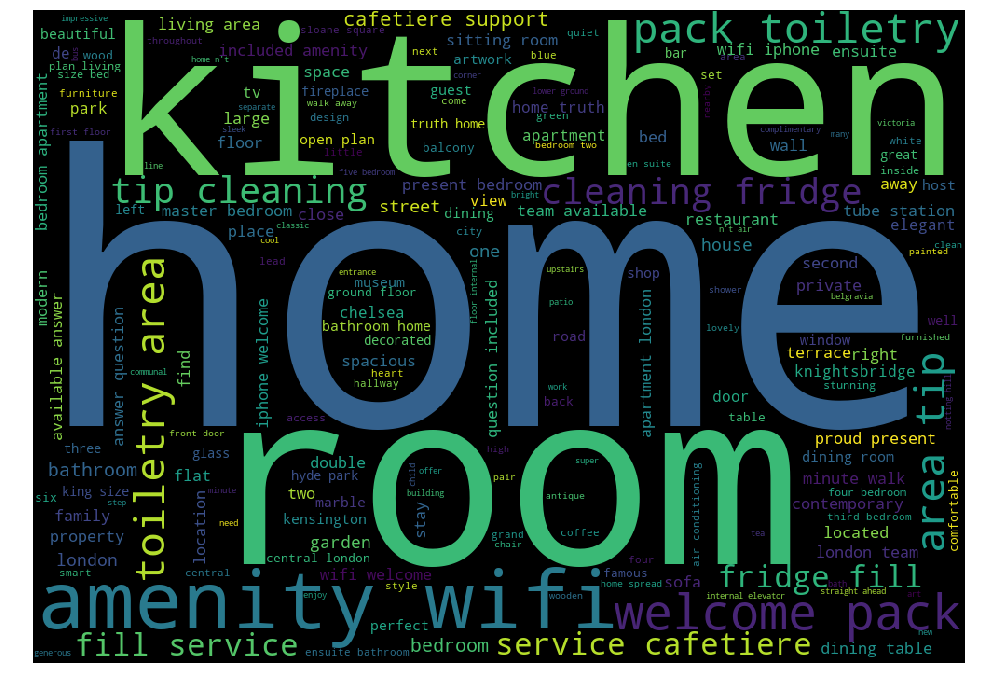
Conclusion 4: It can be seen that unique words like home, south Boston, bedroom, floor, kitchen, resstaurant, spacious, neighborhood, located are very commonly words used when hosts are describing their homes. This is making sense because the primary purpose of airbnb is not to provide luxury hotel suites but just a convinient place to stay. Naturally hosts understand these purposes and create their summaries based on location and requirements in order to attract as many travellers as they can. So if hosts are not able to attract too many travellers then they can add these keywords to their summaries in order to attract travellers to choose their listings

## longitude + latitude

## space



## description

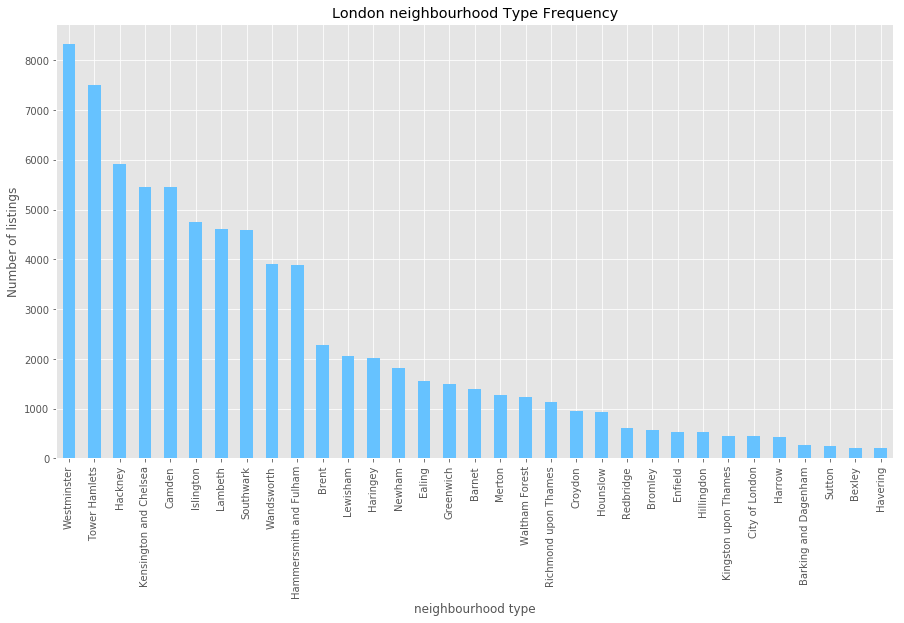


## instant\_bookable

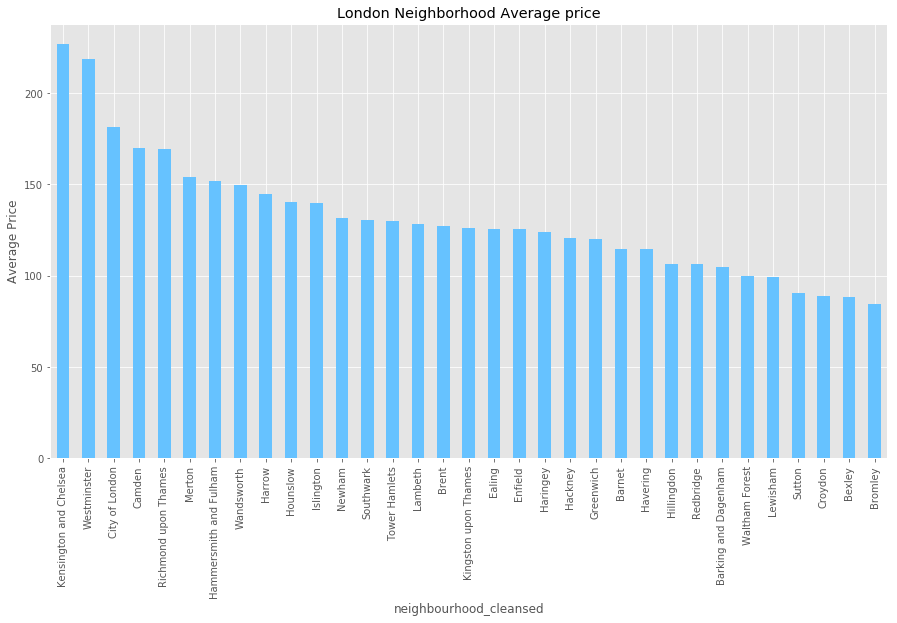
## neighborhood\_overview



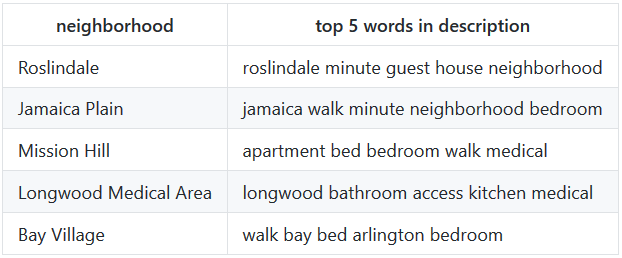
## neighbourhood\_cleansed



=> It can be noticed that the maximum number of listings for entire apartment for Boston on Airbnb is in South End area followed by back bay and Jamaica plain.So these area become potential neighbourhood to invest a property



1. For next analysis I have compiled a list of keywords for differnt neighborhoods in Boston which should be included while writing description about the listing on Airbnb

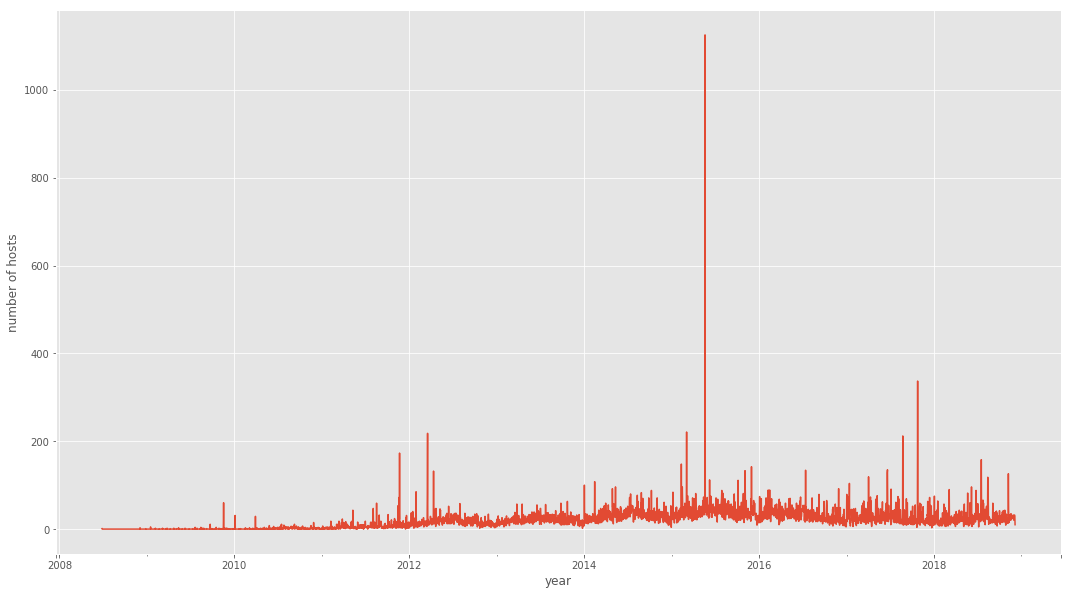


=> I can be observed that for neighborhood "South Boston Waterfront" major words include seaport, "Hyde Park" includes quiet, "East Boston" includes airport, downtown, station and chinatown includes theatre.All these observations shows that these words are unique to a particular neighborhood which gives us another data point that while listing the property on airbnb these words should definitely be included to attract travellers.

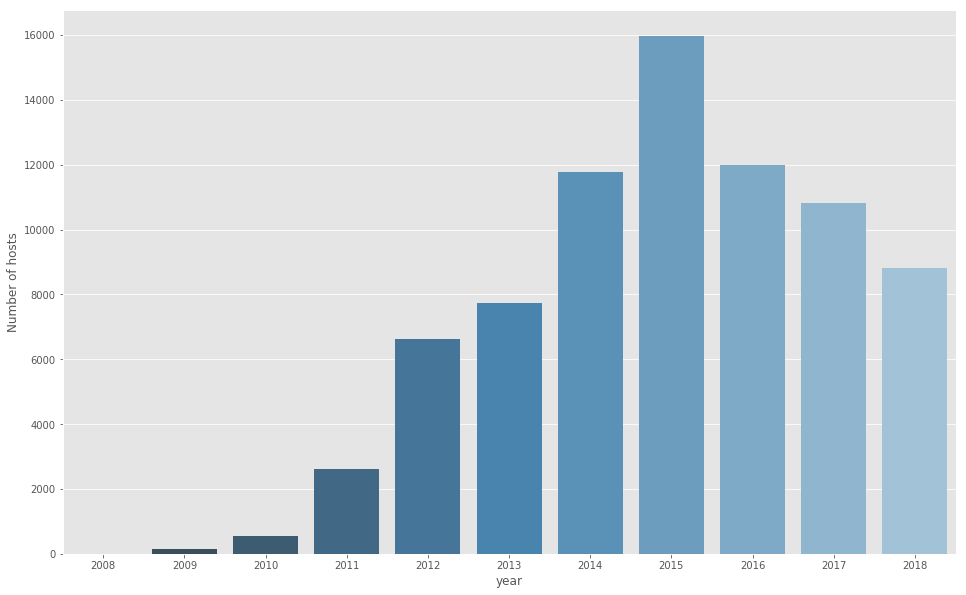
## host\_id + host\_name

## host\_since

1. Lets analyze the trend of people hosting AirBnBs in Boston.We can get insights about how much more popular is AirBnB among hosts now than it was two or three years ago. To analyze the same I cleaned the data and separated the host\_since date in various columns such as day, date and year and plotted a graph for number of host registered vs each day of the year

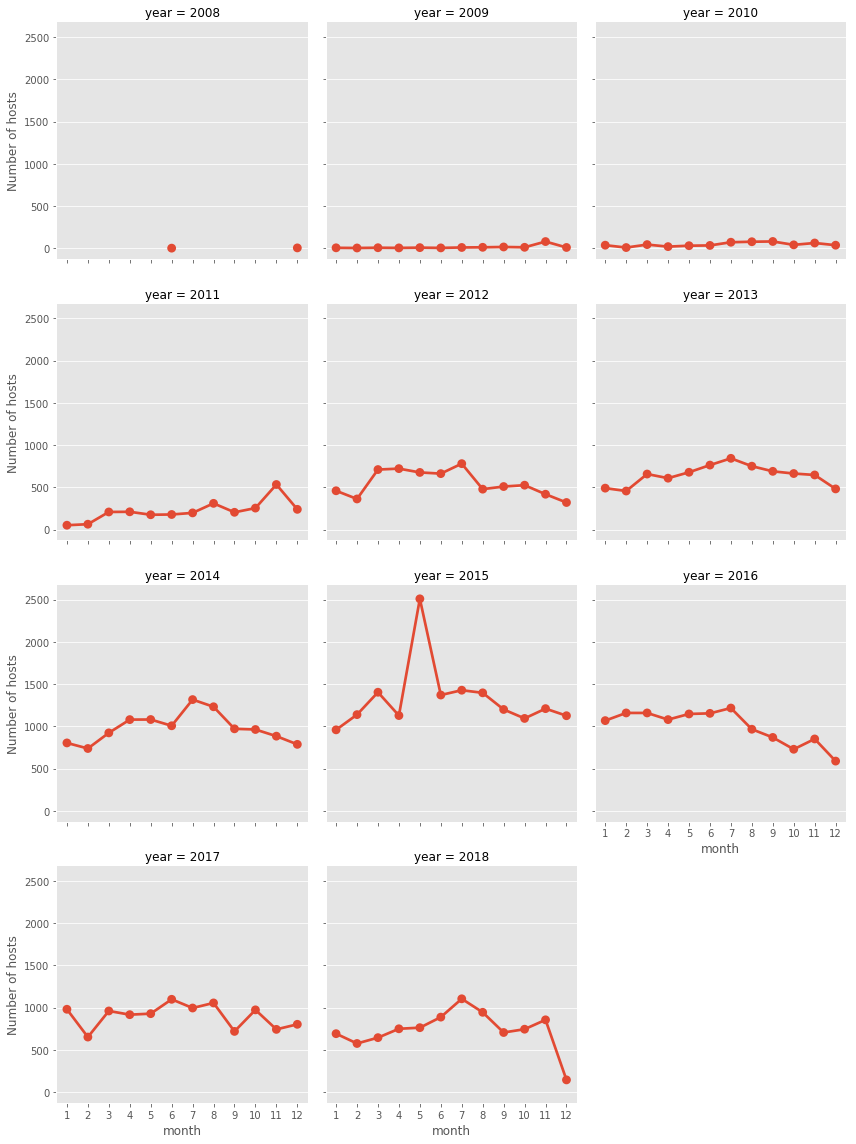


1. I have plotted the same data for each year



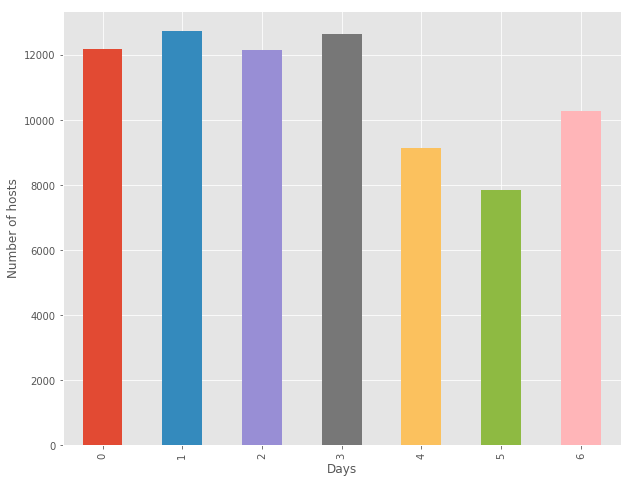
=> It can be analyzed that Most of the hosts registered in the year 2015.

1. Lets analyze the months

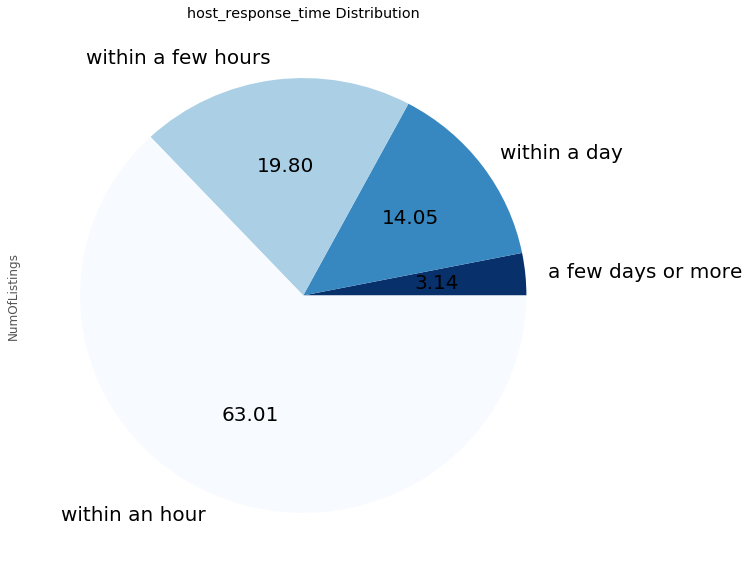


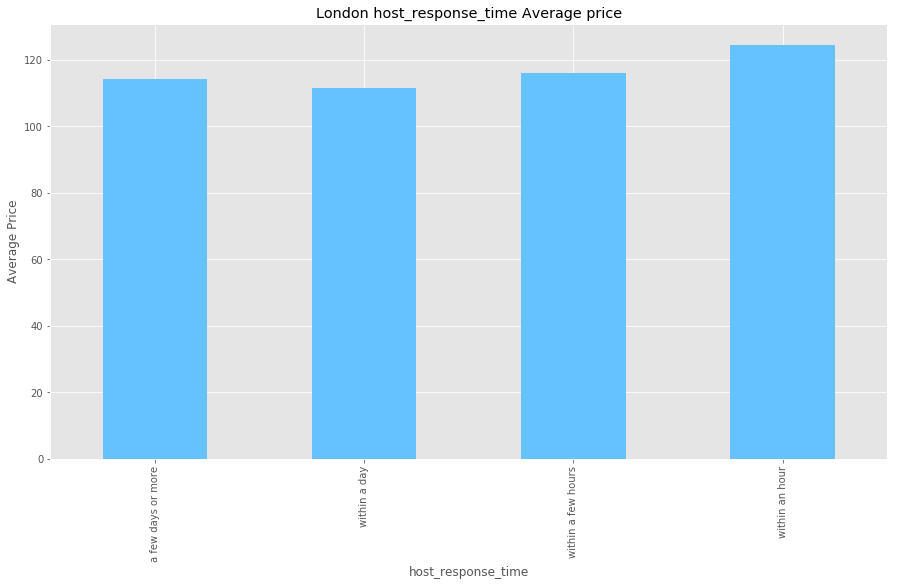
=> it can be seen that most of the hosts registered in the month of july,november and march.Although there is no specific trend followed.Lets check do people register more on weekends?

1. I have plotted the same where 0 corresponds for mondays and 6 corresponds to Sunday



## host\_response\_time





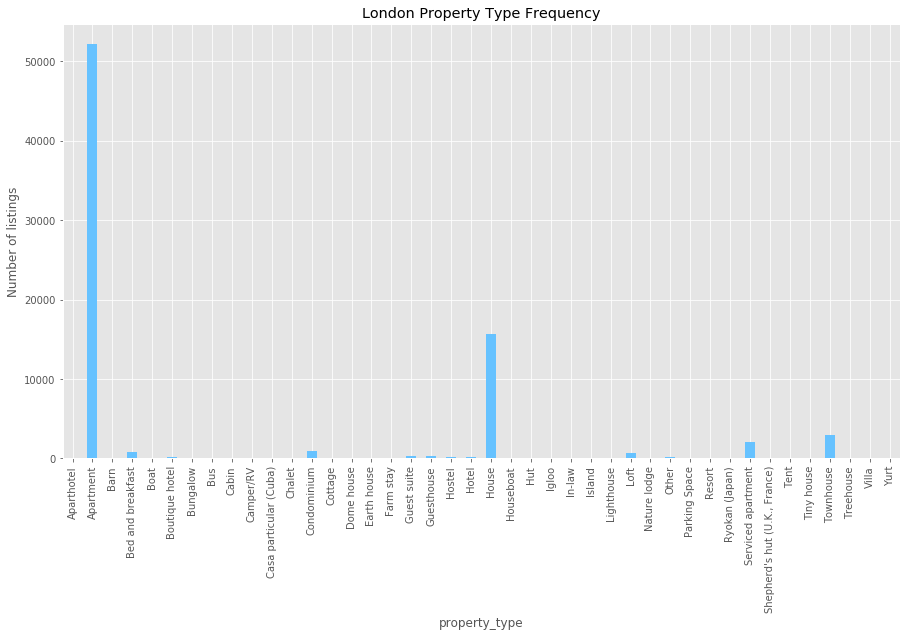
## street

## zipcode

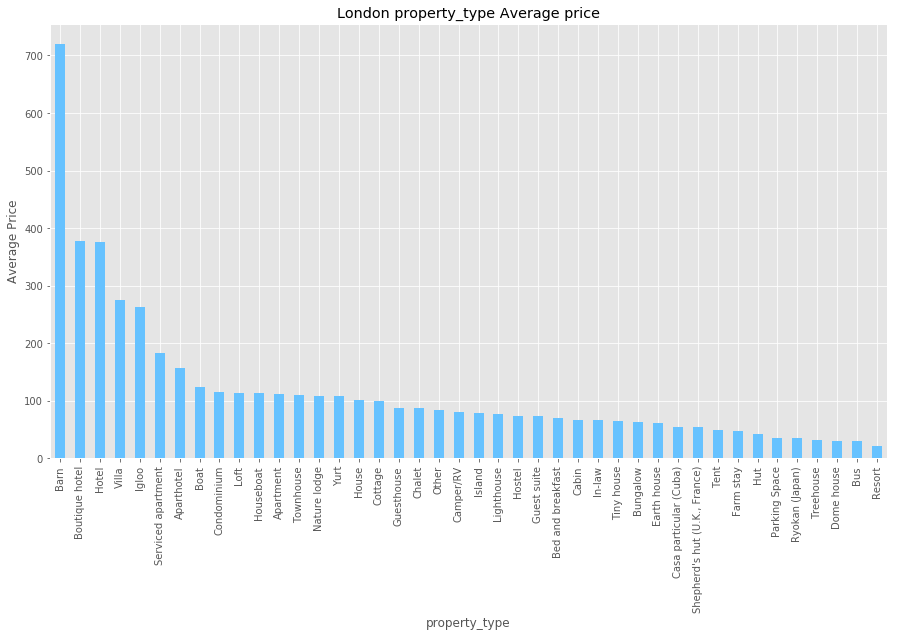
## review\_scores\_rating

## property\_type

Categorizing different listings based upon their property type



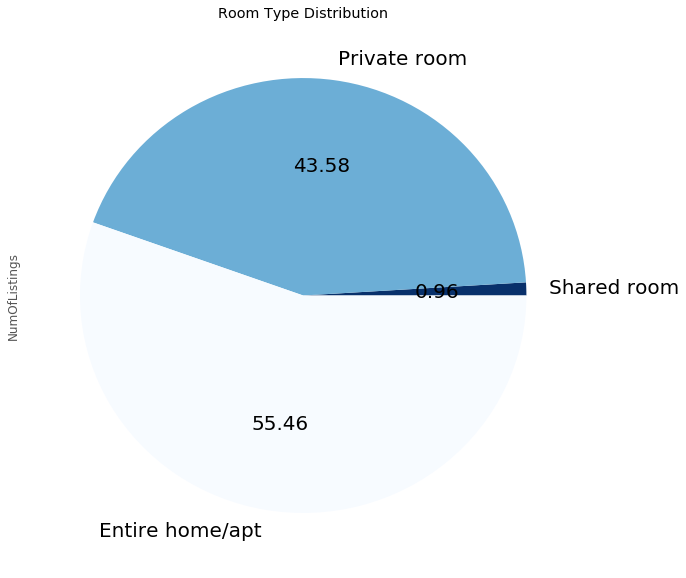
=> It can also be seen that property type also plays an important factor. Not surprisingly, Aparment and houses take up overwhelming majority of all listing, although we do see few instances unfamiliar residencies here and there



## room\_type

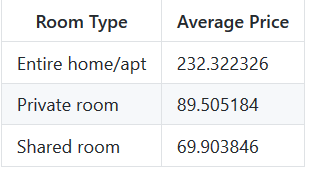
Categorizing different listings based upon their room type

|  |  |  |  |
| --- | --- | --- | --- |
|  | **room\_type** | **NumOfListings** | **Avg Price** |
| 0 | Shared room | 743 | 46.647376 |
| 1 | Private room | 33581 | 56.059557 |
| 2 | Entire home/apt | 42739 | 157.082688 |



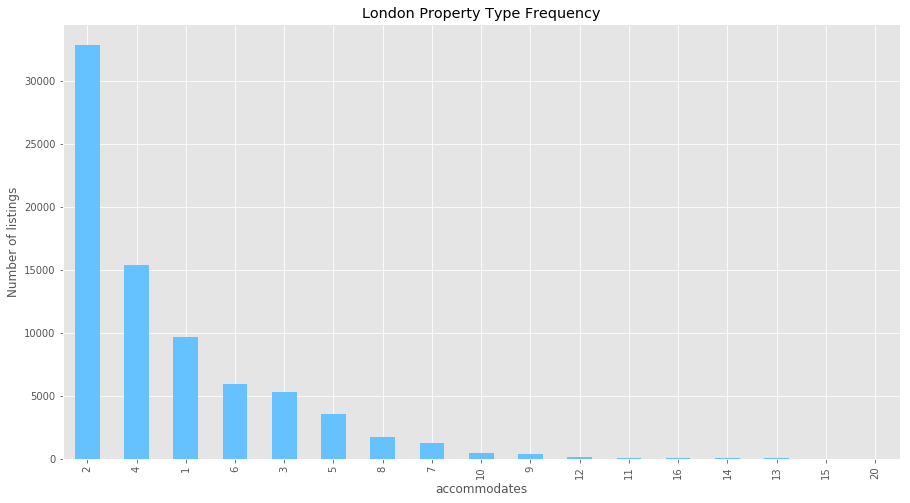
=> It can be concluded that people are more inclined towards listing their entire propery than that of private rooms or shared rooms.

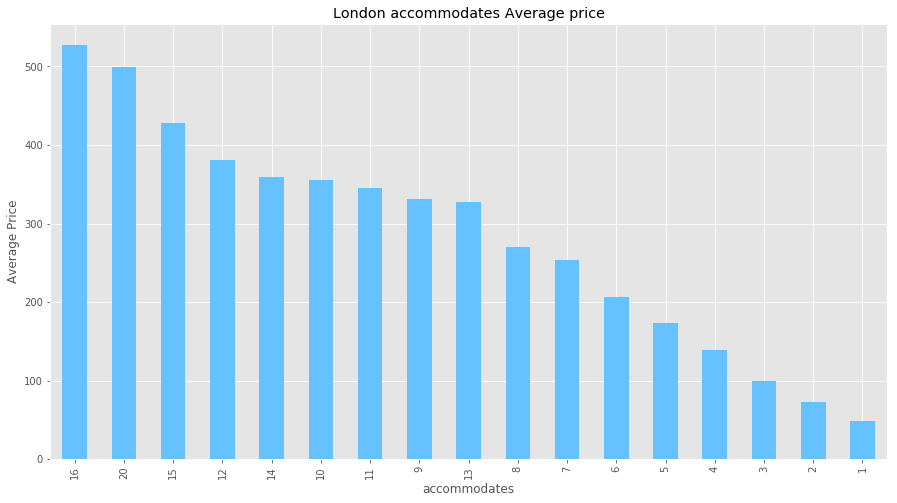
1. It has been analyzed earlier that the maximum number of listings are for Entire Home/Apartment. Lets check average prices for these listings based on room type



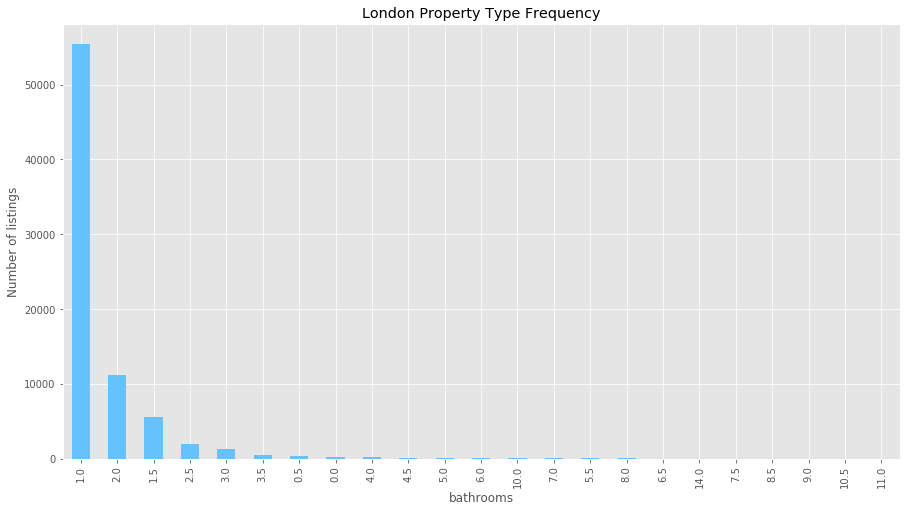
=> It clearly shows that Average Price of Entire home/Apartment is very high when compared to Private room and shared room.It can be concluded that Entire home/Apartment room type has the maximum average price which gives us a data point that after buying an Apartment, listing it as an entire Apartment on Airbnb will help to generate maximum revenue.

## Accommodates

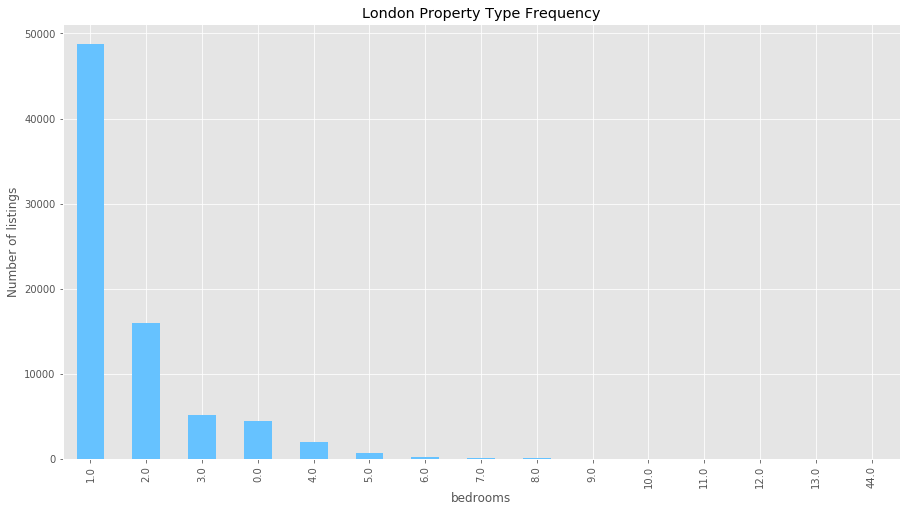




## Bathrooms



## Bedrooms

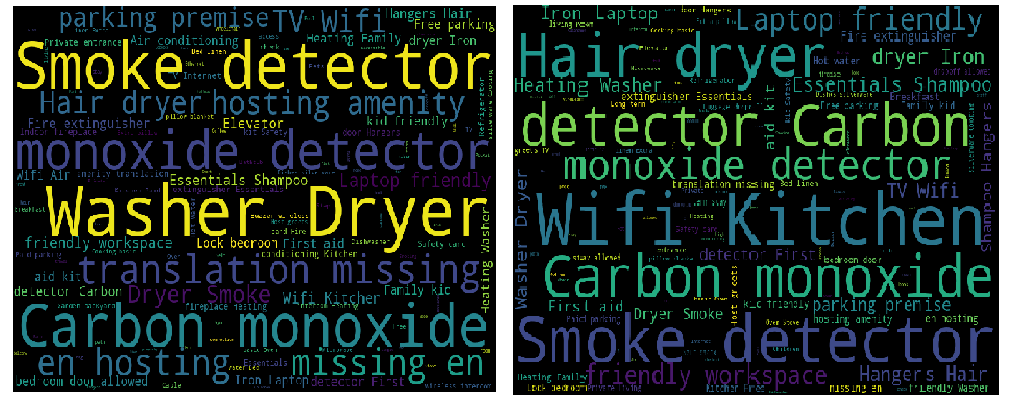


## beds

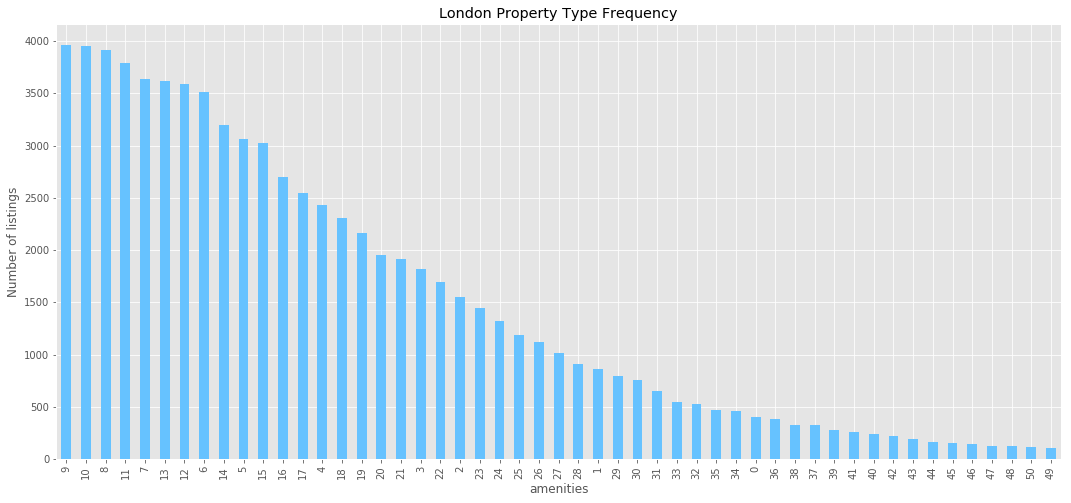
## reviews\_per\_month

## amenities

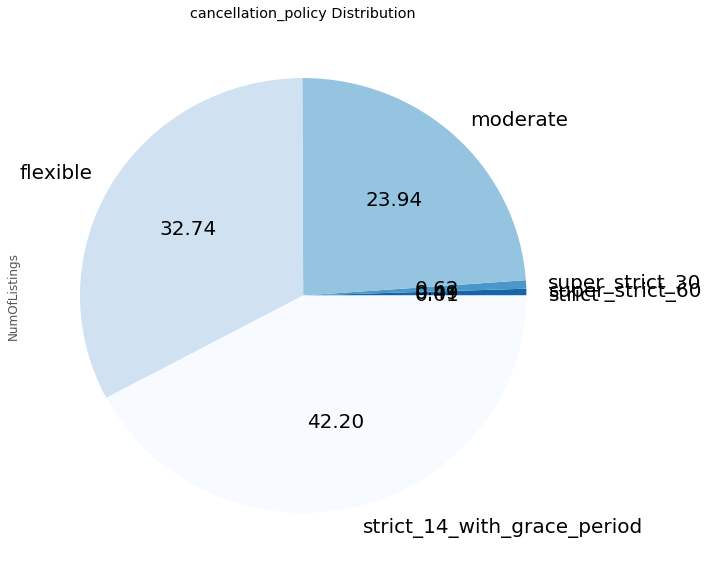
1. Now trying to anlyze how the amenities provided by the listing is related to the price of the same. Two different wordclouds are plotted to help to find out what extra emenities are provided by listings with higher price

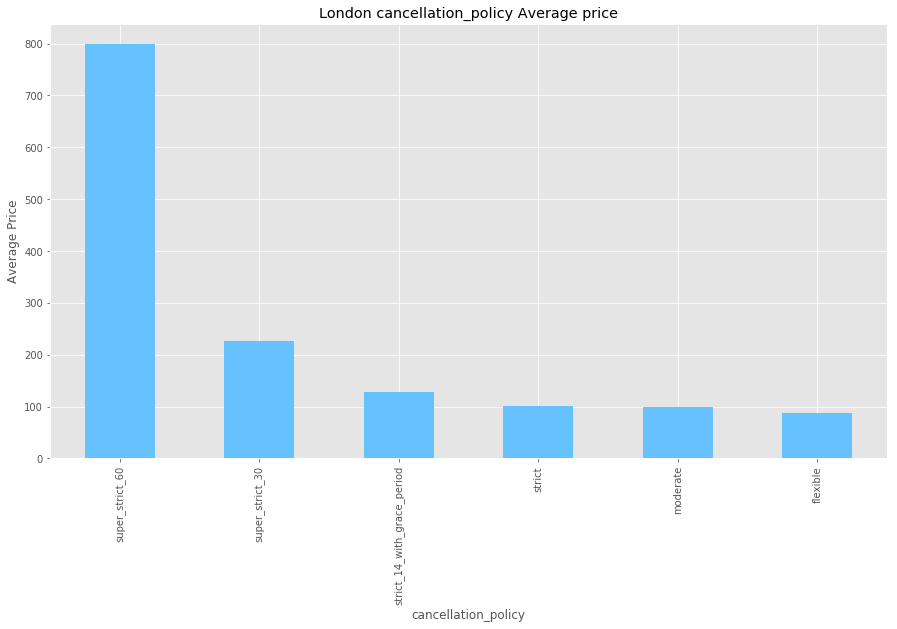


=> It can be clearly seen that listings with higher prices have extra eminities such as Air conditioning, washer/dryer, Kid friendly, Heating, hair dryer, buzzer and Extra emenities comes with extra prices



## cancellation\_policy

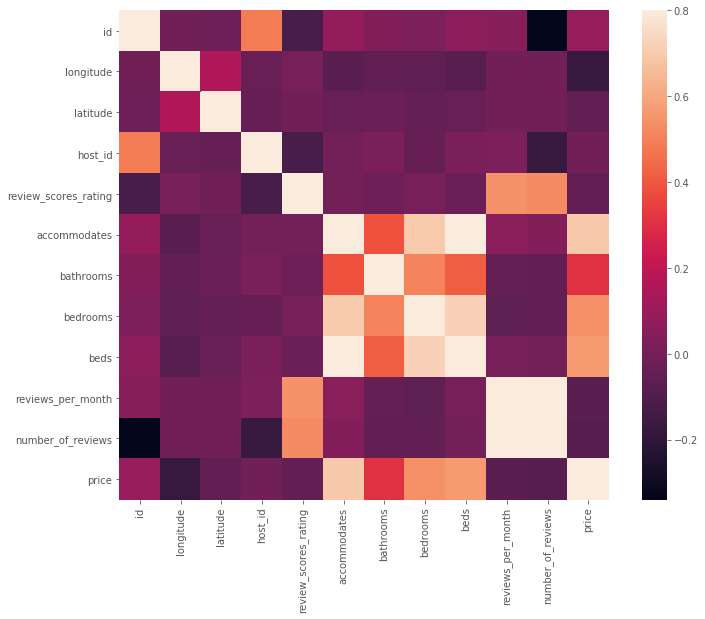




## number\_of\_reviews

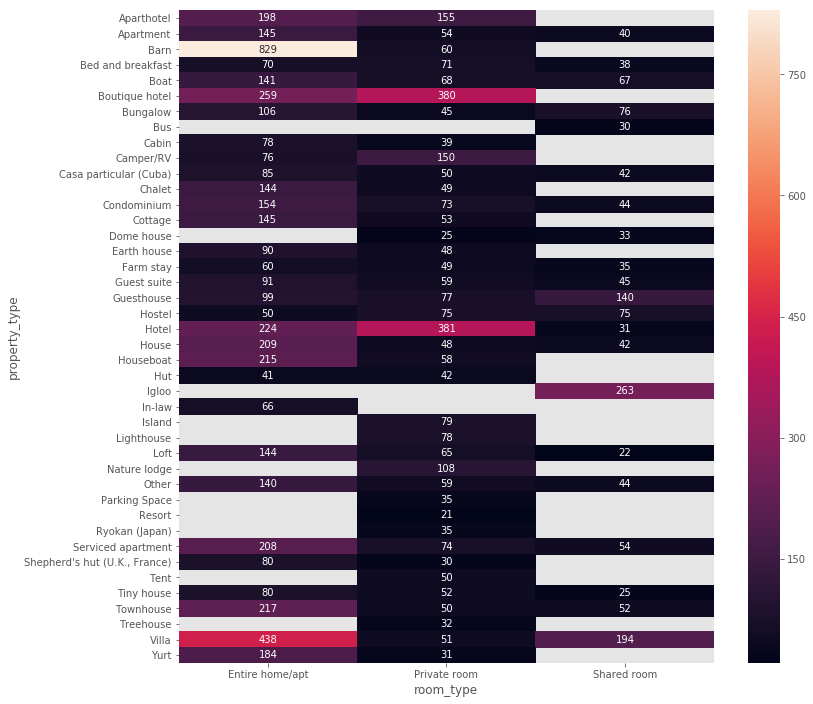
## price

### correlation



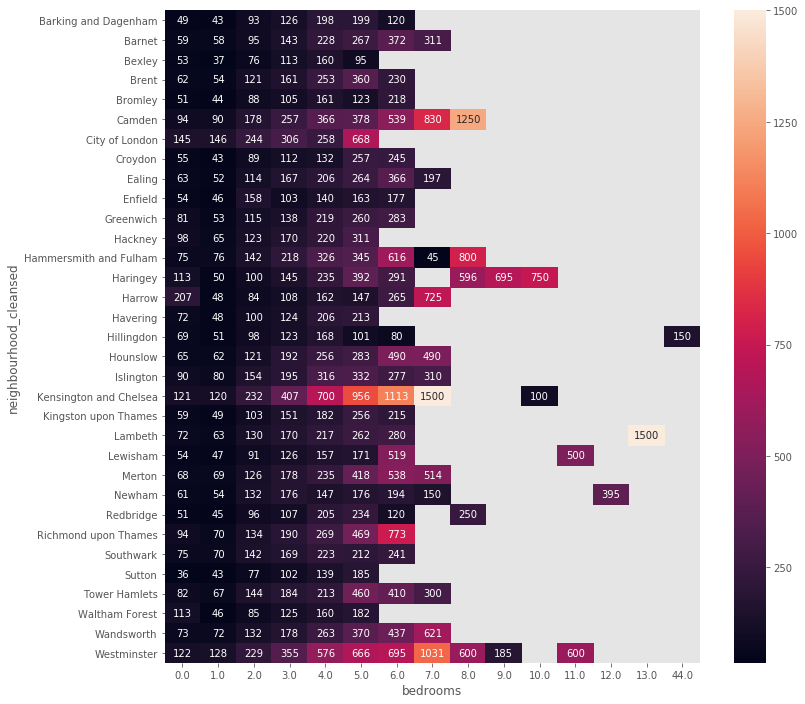
### What causes difference in prices of listings?

1. Analyzing the prices for different room type and property type for the same which includes category and subcategory for all those categories with mean price for each of them



=> This chart allows us to see all the listing’s prices broken down by property type and room type. This gives us a much better understanding of the price broken down in Boston based on property and room types. It can be analyzed that for almost all property type, prices for Entire home/apartment is the maximum. This tells us that Property type and room type plays a very important role in deciding price of listing

1. Show the number of bedrooms available affects the price of listing



Conclusion 3: It can be analyzed that with the increase in the number of bedrooms price of listing increases. Although, it depends upon the neighbourhood as well

So the analysis gives us data points that the prices of listing on Airbnb depends upon the room type, property type, number of bedrooms and neighbourhood. It can be seen that the propery with type as Aparment and the listing as with type as entire house with maximum number of bedrooms has highest price. Although it depends upon neighbourhood as well which is analyzed in the next analysis

**Conclusion**: It can be concluded that prices of listings depends upon following factors:

1. The type of room chosen by the traveller and mostly booking an Entire property costs maximum followed by private room and shared apartment
2. The type of property chosen by the traveller and it can be analyzed that Townhouse and houses are the properties with maximum prices and apartments, houses take up an overwhelming majority of all listings.
3. Price of a listing also depends upon the number of bedrooms the property have and the same also depends upon the neighborhood of the property
4. the summary section is the one which helps to attract travellers and analyzed that presence of unique words like home, south Boston, bedroom, floor, kitchen, restaurant, spacious, neighborhood ,located words tends to attract more travellers
5. with the increase in prices the eminities provided by host also increases.

# Time Series Analysis

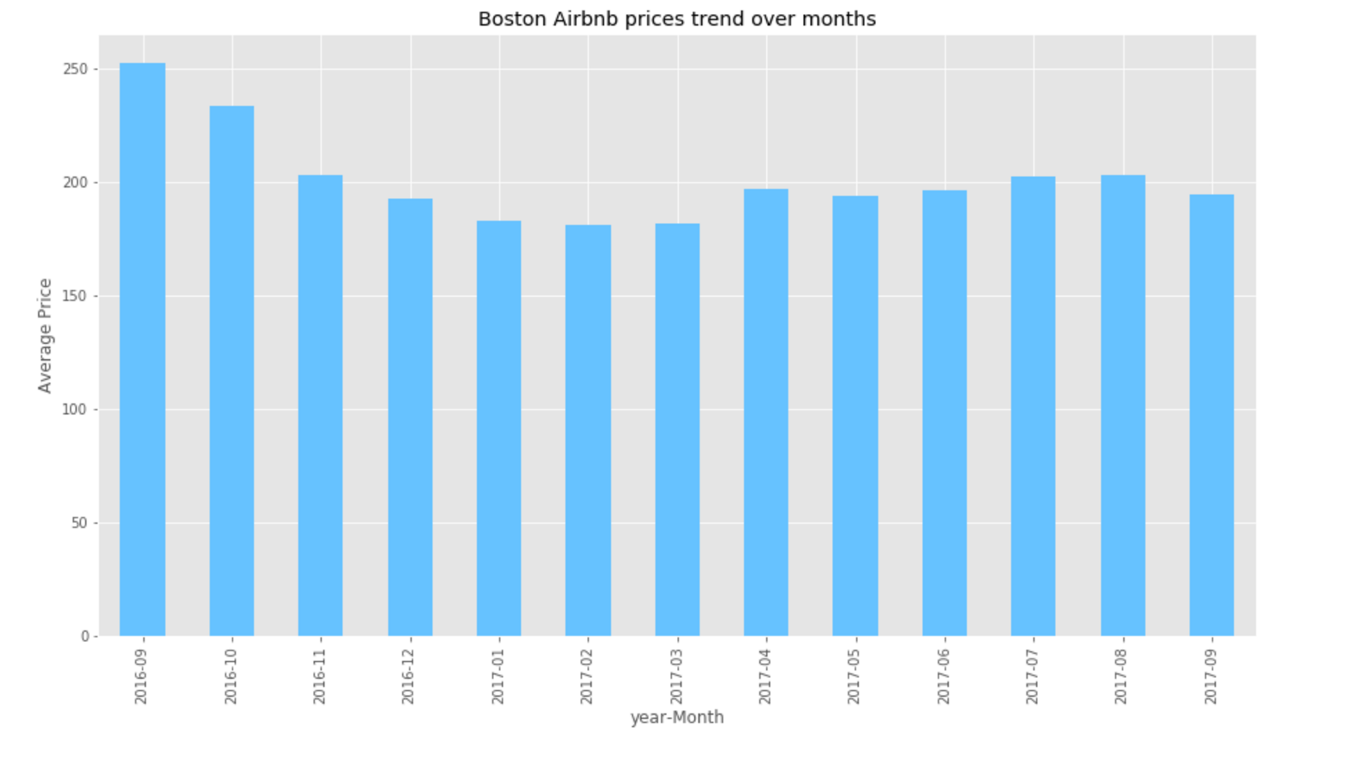
1. What are airbnb NYC’s price changes over years and months?
2. Seasonal pattern of airbnb housing price

## Seasonal pattern of airbnb housing price

It needs to be analyzed that how prices of listings vary with seasons which constitutes the next analysis in which I tried to find out how prices vary according to season,month,week,day

### Boston

1. Cleaning of data has been done and added new columns namely Year,Month and Day by splitting date.I analyzed the data by group it on the basis of Year and Month to see the trend of prices. It can be seen that the data is available from September 2016 to September 2017 and when average prices are analyzed maximum rates for the listings were in the month of september.Visualizing the same for a better understanding.

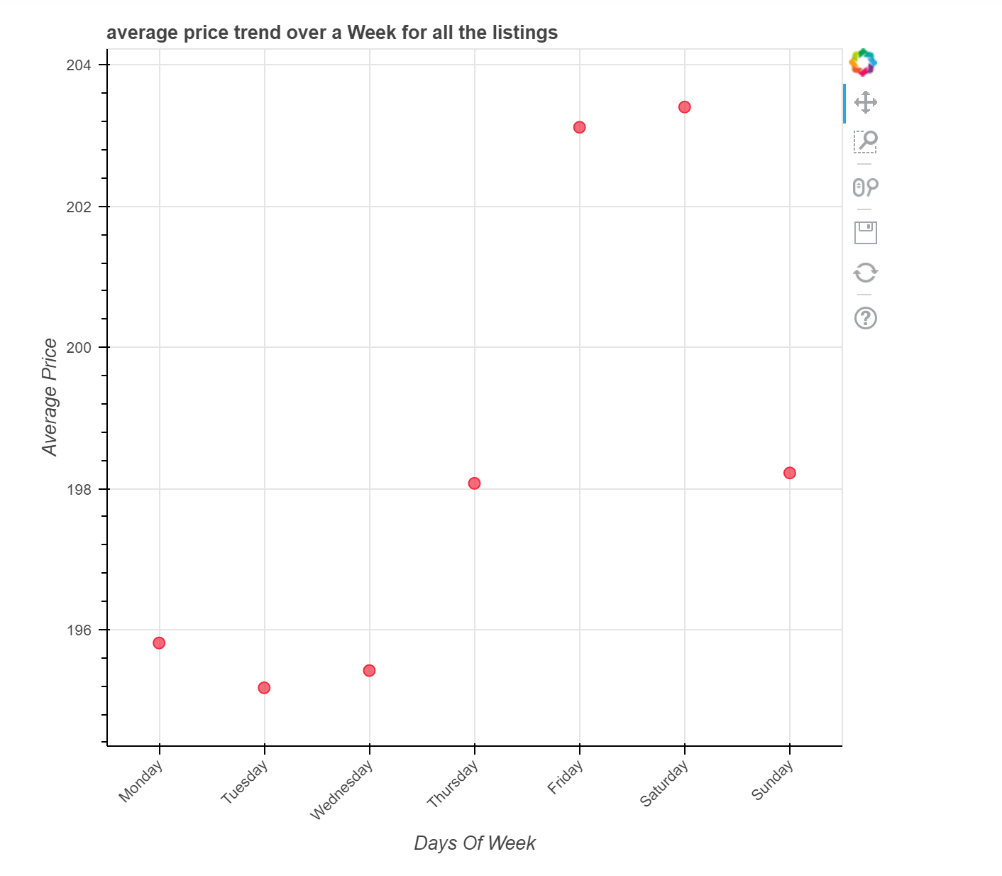


=> It can be clearly seen that the maximum average price for listings were in the month of september and October 2016 and the reason is because of good weather and Massachussetts' best time to observe fall colors.Fall Colors in Massachusetts attracts a lot of visitors which makes September and October peak months for Airbnb hosts

1. To analyze the data further, I have extracted name of the day from given date and checked weather it was a holiday and what is the reason for that holiday using datetime, clendar and holidays modules of python.
2. Added 3 new columns Day\_Name, Holiday and us\_holiday\_name which counsits of name of the day, boolean value for "is it a holiday?" and reason for the holiday respectively. Trying to find a trend of avergae prices for a week

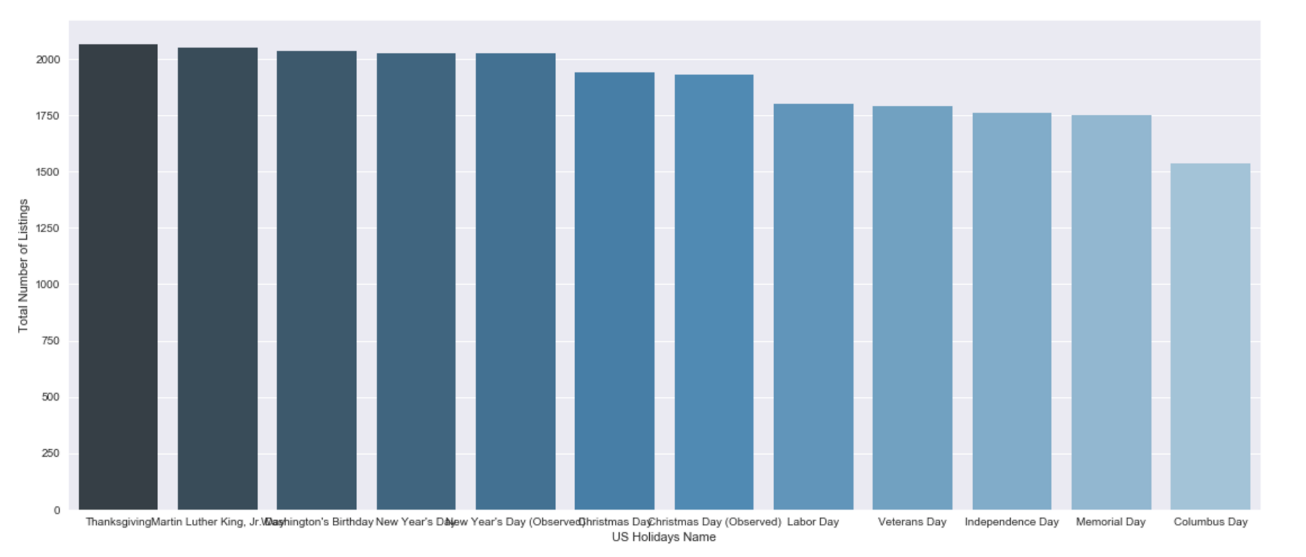


It can be seen that the average price of listings increases on weekends and are usual on weekdays.Lets plot it to get a better understanding



=> It can be seen that the prices are fairly high for the weekends than that of weekdays.

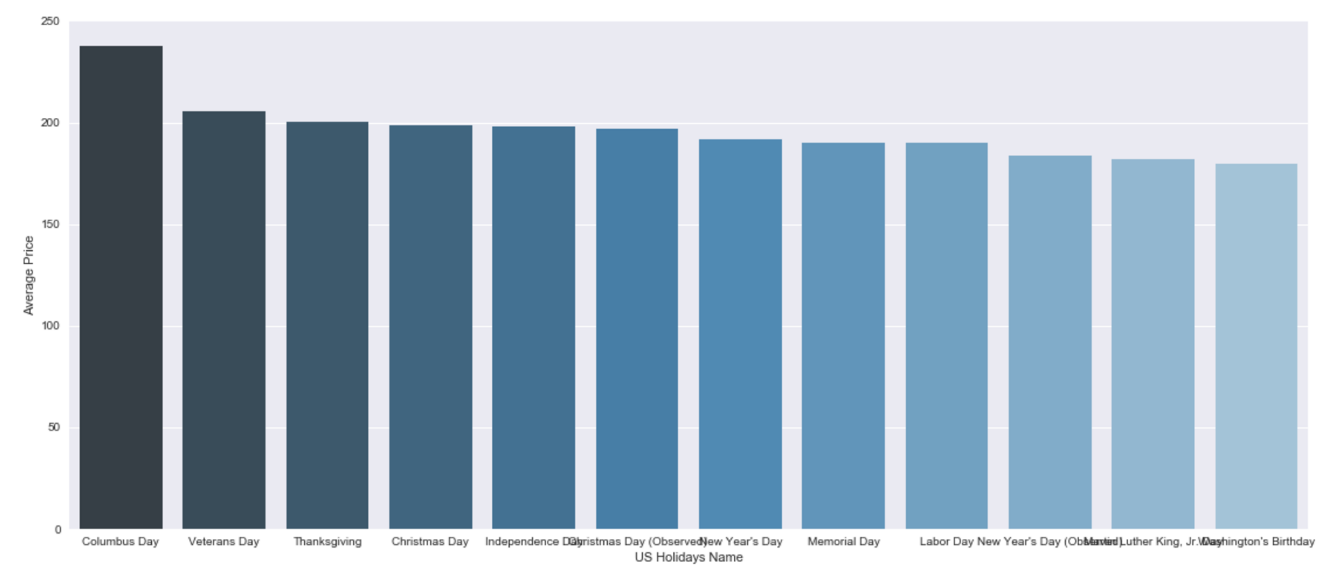
1. Now we need to dig in into Sept 2016 and oct 2016 data to find the reason behind increase in Average prices.



=> It can be seen that the maximum number of listings is for thanksgiving which can be reasoned as its a very popular holiday

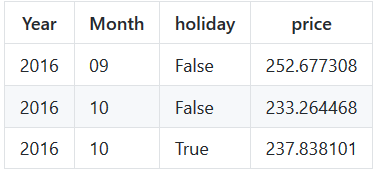
1. .Lets dig in further to find which holiday has the maximum average price





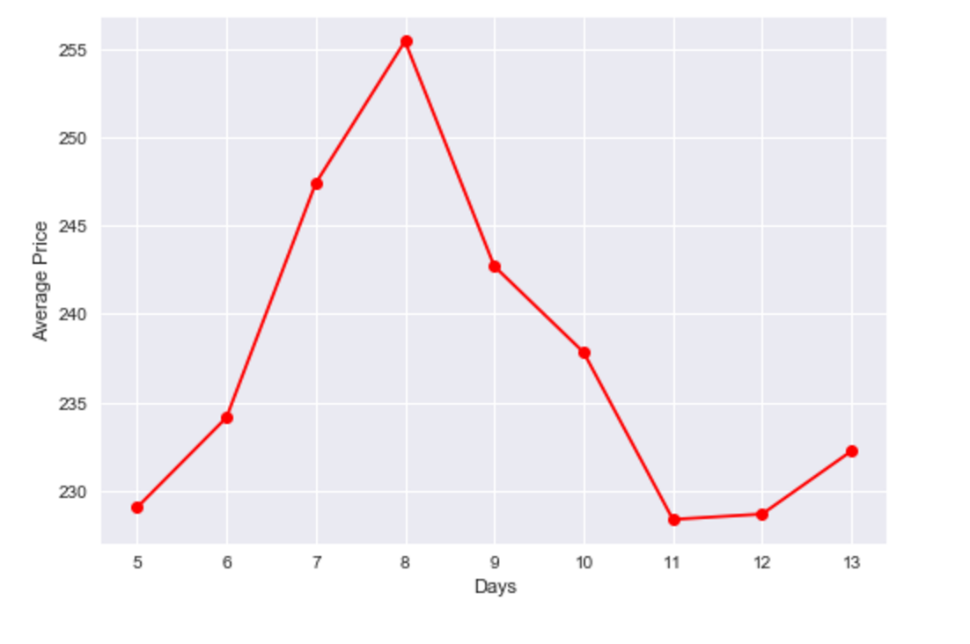
=> It can be observed that the maximum average price of listings is for **columbus day** and that mean Columbus day is the busiest US long weekend holiday for Boston

1. Lets dig in into september 2016 and october 2016 data to find the reason for increase in average prices.



=> It can be analyzed that on a long weekend the average price is 244.63 while for the whole month the average price was 233.Hence, this suggests that the price of listings increses with the presence of long weekend.

1. analyzing data from date 5th of October to date 13th of October which includes both long weekend and normal workdays



=> hence this graph clearly shows that Long weekends have higher prices of listings than that of normal days which gives us another data point that price of a listing increases with the presence of Long weekend.

**Conclusion**: After adding up all the analyis done so far for observing seasonal changes in prices,it can be concluded that:

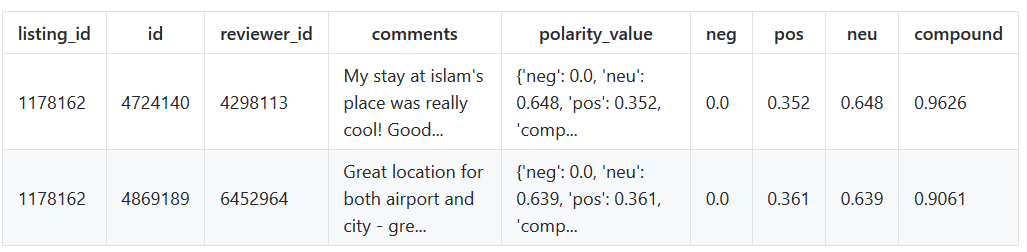
1. If a traveller is visiting Boston on a low budget then he should avoid visiting during September and October as these are the times when average price of listings are fairly high as compared to the other months
2. If the taveller is on low budget then he should avoid visiting Boston on Weekends as prices of listings on weekend are higher than that of weekdays.
3. Prices of listings also depends upon US holidays and longweeknds.Prices on long weekends and US holidays are higher than compared to other days.So for a low budget travel, a traveller should avoid travelling on a long weekend,Specially the long weekend of October i.e columbus day

# NLP Analysis

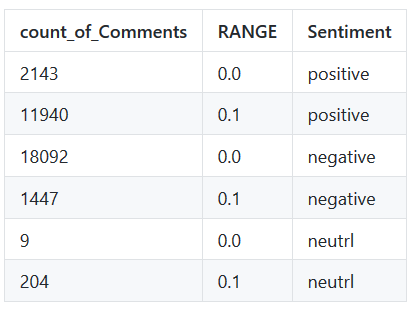
1. Sentiment Analysis & its relation to housing price

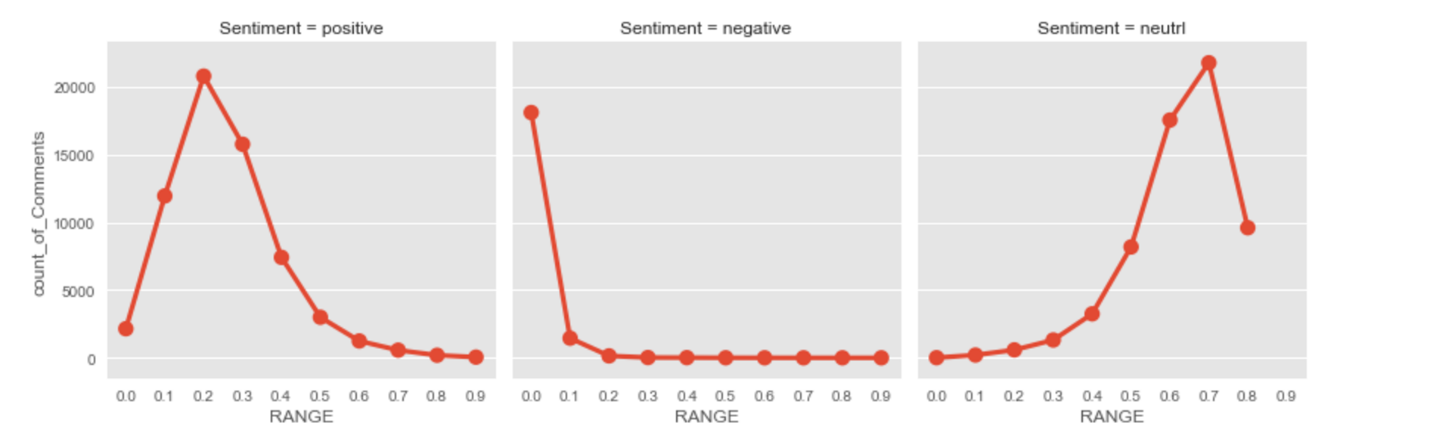
## Sentiment analysis of reviews & its relation with rice

1. There are so many factors which contributes towards the price of a listing on AirBnB.While, we already have few conclusions for relationship between various factors and their dependency on prices of a listing,lets analyze if price of a listing dependent upon number of reviews and if yes, how does it varies? To retrieve the 'sentiment' of comments - 'positive','negative' or 'neutral' I am using built-in analyzer in the NLTK Python library to assign polarity scoore to each comment.I have assigned Ploarity score to every comment which includes detail like Positivity, negativity neutral and compound



1. Our data presented in the dataframe consists of comments in various language while we are concerned only with comments in English language.In the next section I am removing all those rows which consists of comments in different language than that of English using langdetect library.
2. I created graphs to visualize the positive, negative anf neutral score by dividing thernage from 0.0 to 1.0 and counting the reviews between this range. Plotted a factorgraph to undersatnd and compare the sentiments of comments travellers mentioned on the listings





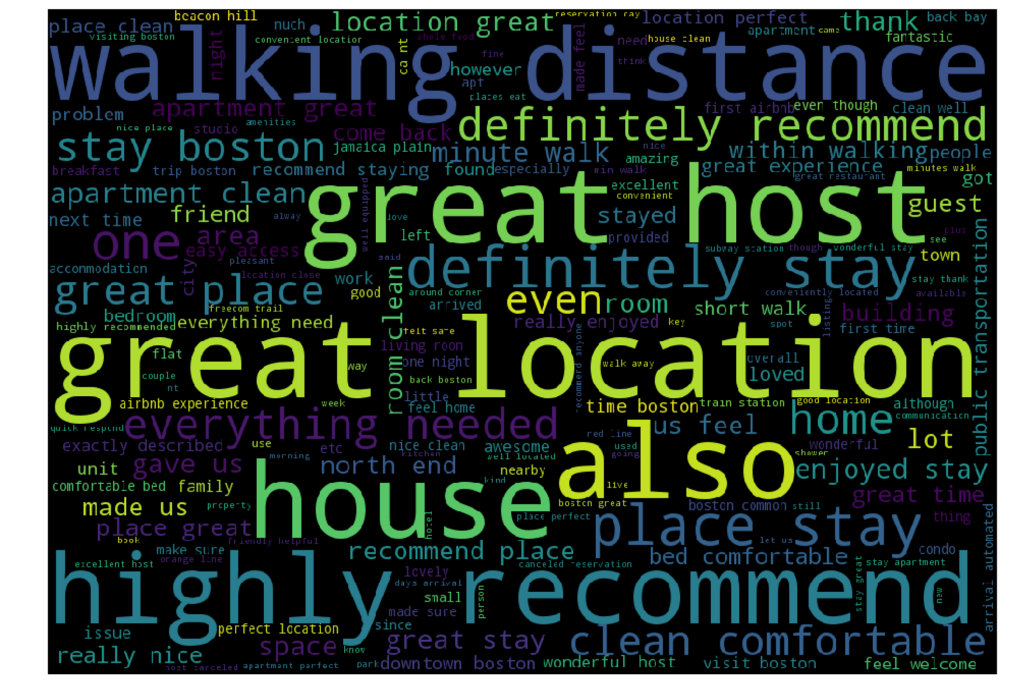
=> It can be seen that Almost none of the texts are classified as having significant amounts of negativity. In fact, a significant amount of them are given exactly 0.0 negativity.It can be clearly seen that most of the comments are neutral.However, a signifigant amount of comments are positive.

1. We can loosely interpret number of reviews as times people have stayed in the said listing. Of course, this depends on when the listing appeared, how long it appeared for, and various other factors. But it could serve to be useful information to see correlation between price and number of reviews.Lets check if there is any relationship between number of reviews and price of listing



=> The graph shows that listings with prices that range around 100 - 400 get the most reviews, probably because they are in the most reasonable price range. The number quickly declines as the price goes up. This indicates that more people book listings that are around $100 - 400 in prices. This shows that it is not necessary for an expensive listing to have large number of reviews.Heneforth there is no exact relation between Prices and Number of Reviews for a listing.

1. Lets analyze what were the most talked about words in all the comments.



=> It can be seen that the most talked about words are "great location", "great host", "walking distance" and "highly recommended".All these reviews and comments plays a big role in attracting the attention of travellers and if there are comments such as "highly recommended" then travellers surely takes a look at the listing

**Conclusion**:

1. It can be clearly seen that most of the comments are neutral and a signifigant amount of comments are positive and almost none of the texts are classified as having significant amounts of negativity
2. Its clearly visible that listings with prices that range around 100 - 400 get the most reviews, probably because they are in the most reasonable price range. The number quickly declines as the price goes up
3. the most talked about words are "great location", "great host", "walking distance" and "highly recommended".All these reviews and comments plays a big role in attracting the attention of travellers

# Invesment/ Other Business opportunities

1. Investment: Where to invest a property to get maximize return from airbnb
2. Predict destination of customer, find the new opportunity (high demand place)
3. What is the overall location distribution of airbnb
4. Factors that influence the trends at airbnb
5. Which boroughs & neighborhoods are recommended for investment?
6. What can be expected in terms of annual income?
7. Airbnb’s impact of housing
8. Hotels & Airbnb
9. Should I airbnb my property

# Airbnb

1. Price recommendation engine for airbnb
2. Analyzing trust factor
3. What airbnb revenue can tell us
4. What kinds of information hosts provide in other providers to reduce uncertainty and signal trustworthies
5. Personalized recommendation system for airbnb

# Predictive Models

# Others

1. Racist problem

# Reference

1. <https://github.com/ruchigupta19/Boston-Airbnb-data-analysis>