Airbnb- Zillow New York Data Analysis

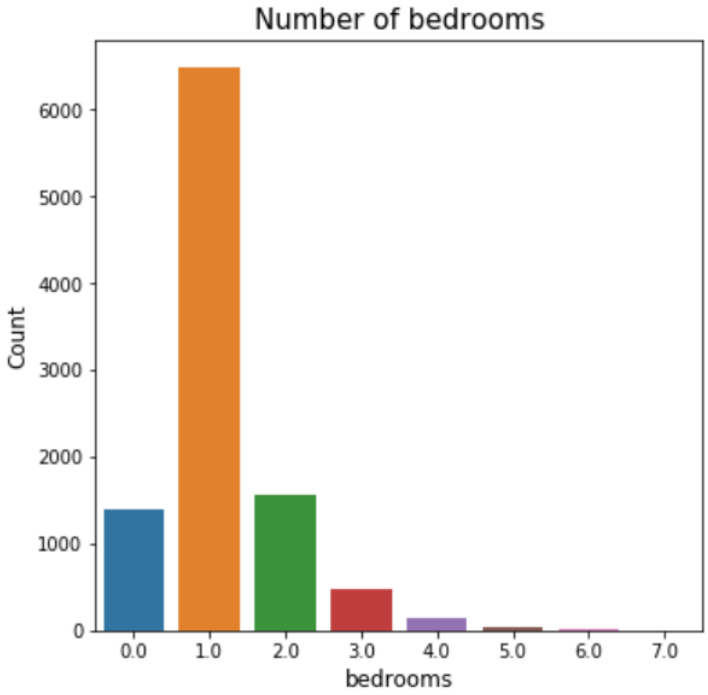
A real estate company that has a niche in purchasing properties to rent out short-term as part of their business model specifically within New York City wants to understand which zip codes generate the most profit on short term rentals within New York City. The company has already concluded that two bedroom properties are the most profitable.

In this report I will analyze Airbnb data for New York's 2 bed-room properties using Zillow's median home value dataset that has price values from 1996 to 2017. An occupancy rate of 75% is assumed for the calculations.

The analysis is done on Jupyter notebook using libraries such as matplotlib, seaborn, plotly, pandas and numpy. The insights are described below and the code file and readme file is included in the zip file.

Step 1: Quality Check

- *The two datasets were merged on column "zipcode" for state New York and was filtered for two-bedroom properties.*



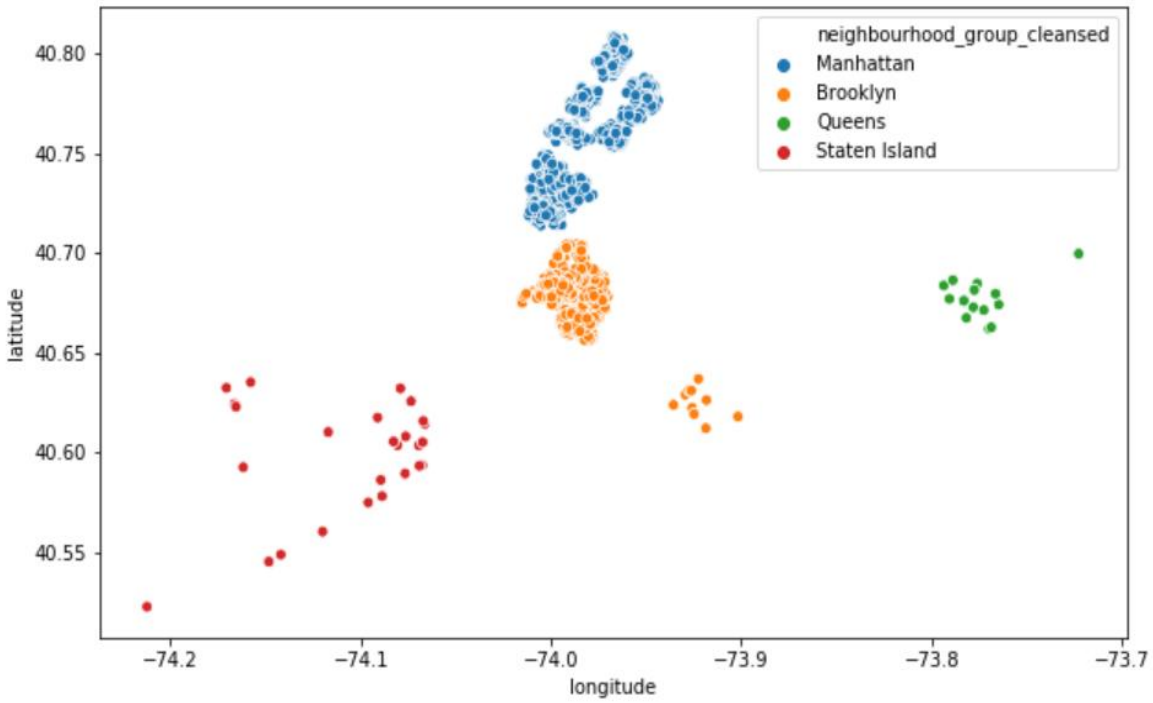
Majority of the data is listed for 1-bedroom properties but since it was concluded 2-bedroom properties are more profitable we will consider only these properties for our analysis.

- *There were some zip codes present in incorrect format such as 11385-2308, 11249\n11249 that were removed from Airbnb dataset to perform smooth merging of the datasets.*
- *Zipcode 10013 is present in both Manhattan and Brooklyn but originally, it's in Manhattan. So, Brooklyn zipcode was dropped from the dataset.*

- ***Missing Data:***
  Columns with more than 95% missing data were dropped as any mean/median imputations in them wouldn't have been reliable for analysis.
  The analysis done on the variables below had majority of the data present and hence the analysis is reliable and accurate.

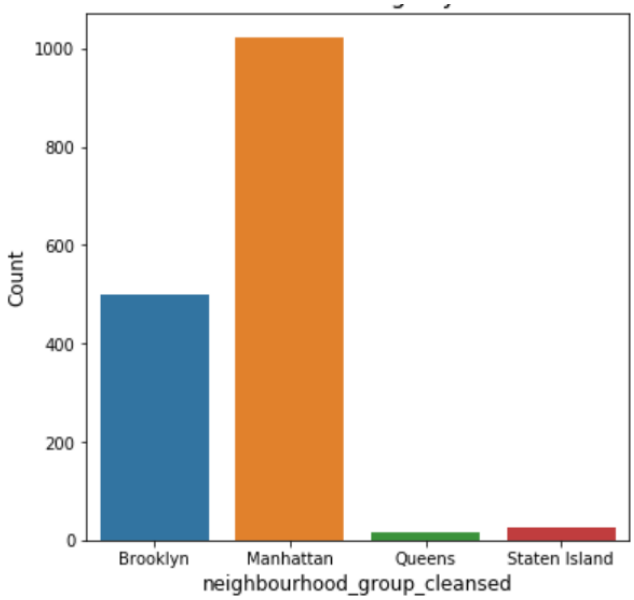
Data Insights:

- *Number of Listings per neighborhood group:*

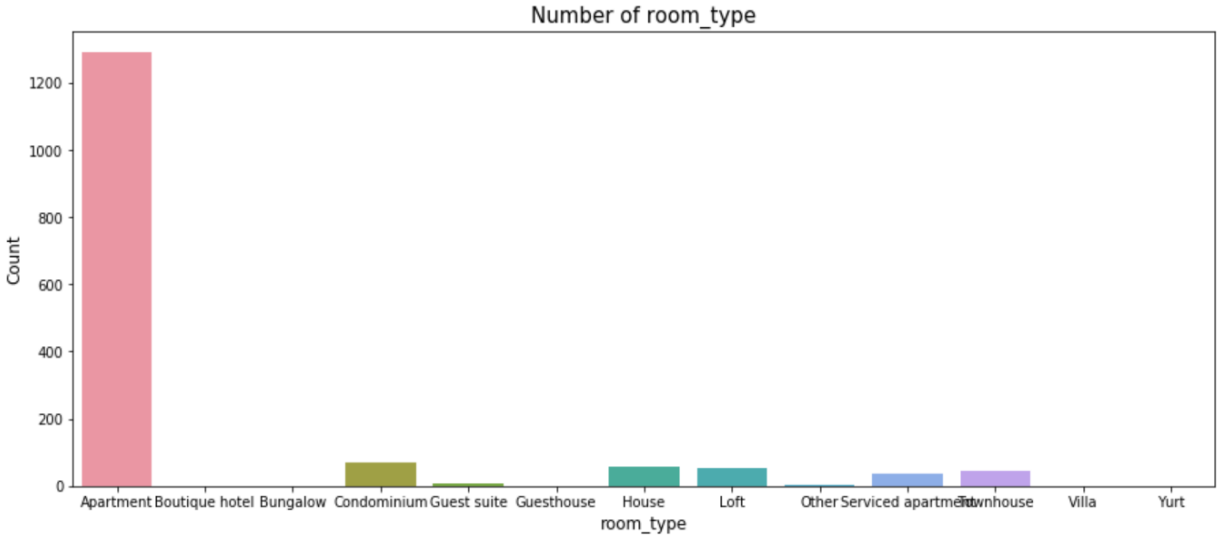


The plot above shows that Manhattan has maximum number of listings followed by Brooklyn, Staten island and queens.

Count of listings per neighborhood group:



- *The graphs below show distribution of property types across various zip codes, it is evident that Apartments type of property is the more popular one than any other property type.*

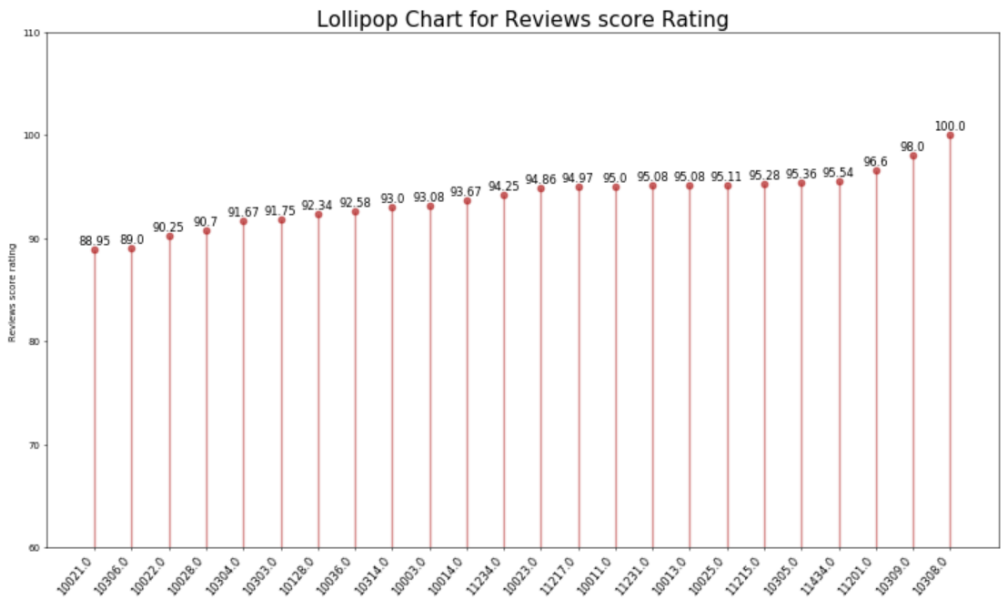


- ***2017 mean house price value for zipcodes:***

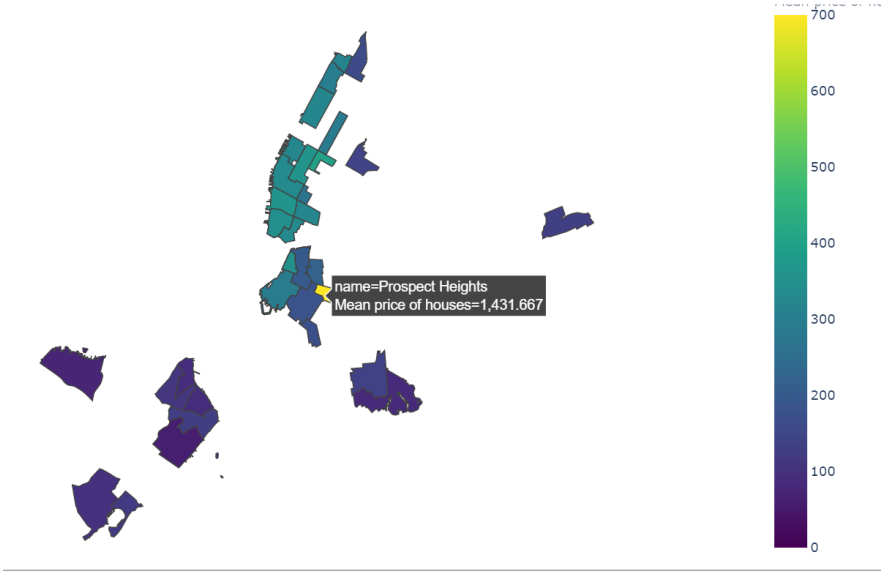
From the above plot it can be seen Manhattan zipcodes (10013,10011,10003) has the highest property values as compared to other neighborhood groups.

- ***Review score Rating***

zipcodes 11201,11215,10025,10013 and 11231 have the highest ratings above 95 with count of 60 and above, hence are reliable



- ***Price Per night for neighborhoods***



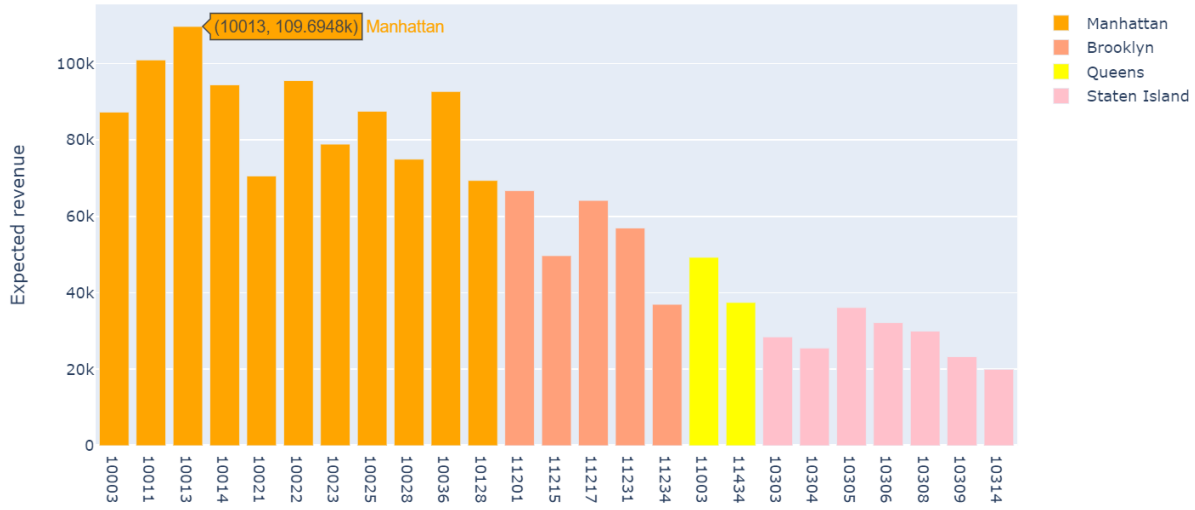
From the above choropleth we can see that Prospect heights (Brooklyn) has the highest price per night followed by Manhattan neighborhoods. Staten Island has the lowest price per night values.

Final Insights:

In Airbnb dataset the price is mentioned for per night stay whereas Zillow has monthly median value of house prices. So, converting them into same units of time.
For house price I have taken mean value of price for year 2017 from Zillow assuming future values will not differ much from house prices of 2017.

- *Expected Revenue:*

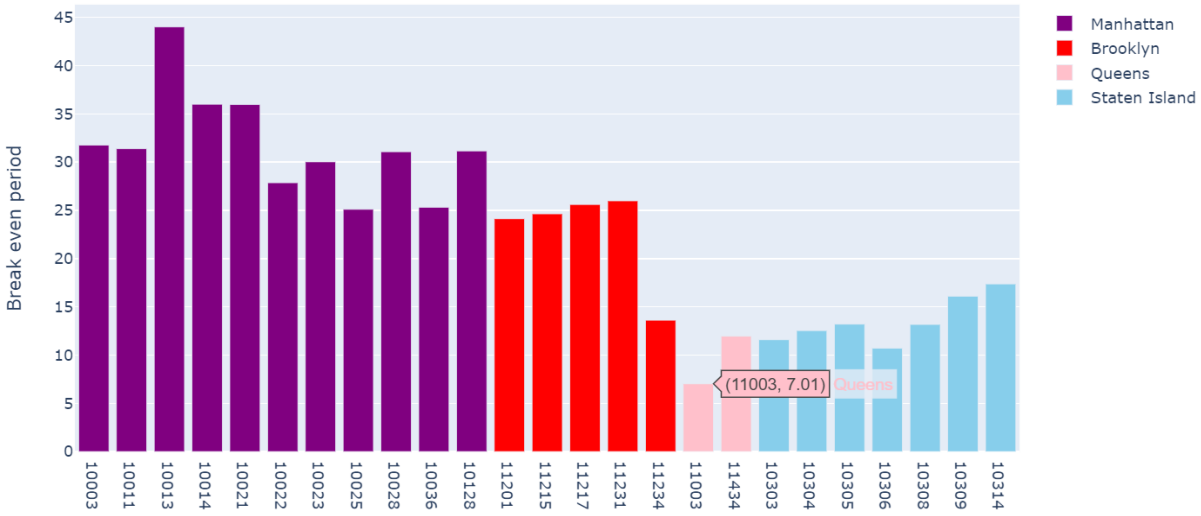
Expected revenue generated by Airbnb will be:
*Price per night * No of days (365)* Occupancy rate (75%, assumed)*



The above interactive plot is built using python's plotly library.
We can see zipcodes 10013,10011,10022,10036 have the highest expected revenue and all belong to Manhattan. This is obvious as above we saw Manhattan has the highest price per night values.
On the other hand, Staten island zip codes 10314,10309,10304 have the lowest expected revenue.

- *Break even Period in years: it's the number of years needed to pay back an initial investment with positive net income.*
  Zipcode with lowest break-even period will be more profitable as the investor will start making profits early.

Break even period= House price/ expected revenue



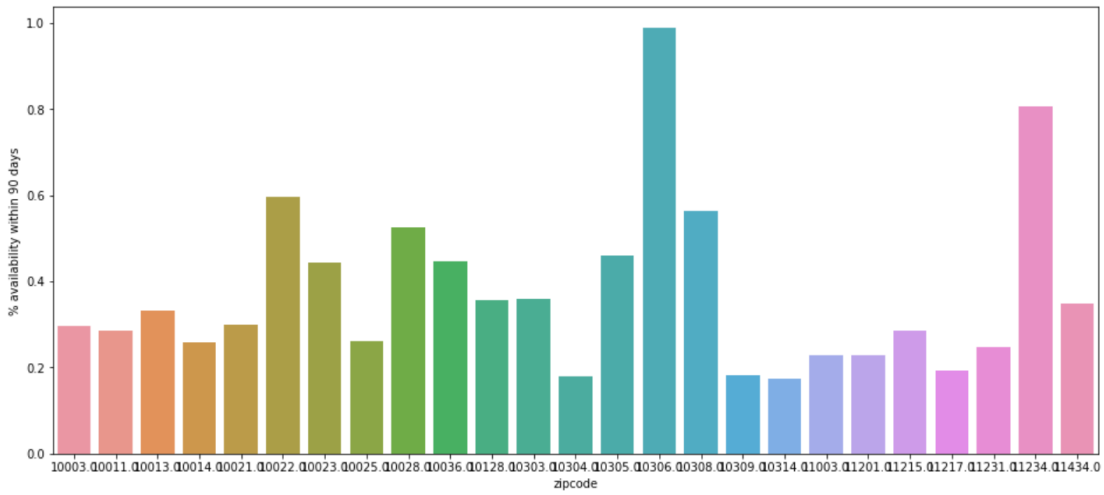
*It can be seen zipcodes **11003, 10306,10303 and 11434** have the least break-even period and hence are profitable zipcodes to invest in.*
We conclude that the above zipcodes from Queens and Staten Island will generate revenue earlier than other neighborhoods.
Even though Manhattan zipcodes has the highest expected revenue their break-even period is quite high and hence not preferable neighborhoods.

- ***Availability of property within zipcodes:***

  Now that we have seen the profitable zipcodes lets dig into the potential zipcodes that can be invested in based on availability of property within 90 days.

  Since we are considering investing in properties for short term rentals, we will see the no. of days properties are available to rent out within zipcodes. Zipcodes where property availability is low can be targeted for good returns.

  Based on the below plot Zipcodes 11217, 10025 and 11231 have availability less than 20% and hence can be targeted as they have pretty good expected revenues as well.



Recommendations:

- *Based on the above analysis and findings, zipcodes 11003, 10306,10303 and 11434 should be considered for investment for short term rentals.*

- *Even though zipcodes in Manhattan has highest expected revenue their break-even period is really high and hence are not preferable.*

- *Based on the property availability analysis, some zipcodes in Manhattan can be considered for investment as property availability is low and can be targeted as they have high expected revenues.*

Step 4: Future Steps

- *Expenses data was not given. There was no other information related to expense data, other than cleaning cost (which was exclusive). Costs like brokerage, taxes were not included. Information on expenses data would have helped in accurate analysis of profitable zipcodes.*
- *SQFT data of the property would have helped in accurately predicting the current purchase price and rent. As more than 95% of the data was missing and any analysis on it wouldn't have been reliable.*
- *Further analysis on factors that impact booking on Airbnb and by what weightage could be developed to get overall rating of the customers.*
- *Model of occupancy could be developed.*
- *Using past data of Airbnb rents future demands could be forecasted so as to make investments in zipcodes where demand will be high in future years.*
- *Sentimental analysis could have been analyzed on reviews of the listings.*