Executive Summary to determine the most profitable zip codes in New York City for short term rentals.

By

Reena Sehitya

06/14/2021

Problem Statement

Our consulting firm has been approached by this real estate company whose core business is to invest in properties, specially within New York city, to rent out for short terms. They already did research and found out that two bedrooms are the most profitable property type within New York city. They want us to determine the zip codes within New York city that would be the most profitable for short-term rentals so that they can make investment decisions in terms of buying property in such areas within NYC.

To do this analysis, we referred to the following publicly available datasets:

- Cost data: an estimate of value for two-bedroom properties.
- Revenue data: The detailed information about all the rental properties of various cities, including New York city, listed on Airbnb.

Assumptions

To conduct the analysis for this task, following assumptions have been made:

- The investor will pay for the property in cash (i.e. no mortgage/interest rate will need to be accounted for).
- The time value of money discount rate is 0% (i.e. $1 today is worth the same 100 years from now).
- All properties and all square feet within each locale can be assumed to be homogeneous (i.e., a 1000 square foot property in a locale such as Bronx or Manhattan generates twice the revenue and costs twice as much as any other 500 square foot property within that same locale.)
- An occupancy rate of 75% for all the properties listed in Airbnb is assumed for the entire year.
- All the properties have the same value and cost for every zip code.
- The Airbnb dataset is for the year 2020 and estimated cost price for June 2020 is considered to be the current cost price.

Key Metrics

To evaluate which zip code will be the most profitable for the short-term rentals, following key metrics were considered in the analysis:

- Cost: Cost is considered as one the key metric to find out how much investment will be required in purchasing a property in each zip code. Higher the cost means higher the investment required to buy the property. Median price of 2020 is the current cost price for performing this analysis. The cost price for 2020 will be calculated based on the available cost data.
- Popularity: Popularity is considered as one the key metric to find out which zip codes are in highest demand. Higher the demand means higher potential for profit. To evaluate zip code on popularity, I have considered total reviews per zip code.
- Annual Rate of Return: This key metric is considered to find out how much rate of annual return is to be expected from each zip code. To calculate this metric, I have considered median house price of 2020 as the total cost of investment and annual rent from that property as the return in investment from that property. Annual rent is computed as the product of nightly price and occupancy rate (i.e. 75% of 365 days in a year).
- Variability in terms of rental price: This metric will be used to measure the potential risk involved with investing in properties for a particular zip code. Potential risk is the fluctuation in the rental price of the properties. Higher the fluctuation, risker will be investment in the property since the revenue and cash flow from the property will not be stable throughout the year.
- Most options: This metric measures the zip codes with the greatest number of rental properties thereby representing the zip code with the most number of options to rent a property for a short duration.

Exploratory Data Analysis

Data cleaning and Processing of the Zillow dataset

Following steps were taken to clean the Zillow dataset:

- Changed the datatype of RegionName from integer to string datatype.

- Defined a function to determine the percent of missing values in each column of the dataset. Then called that function to compute the percent of missing values in each column of the dataset.
- Dropped all the columns with null value(s)
- Created a new dataframe containing only the median price for each zip code in NYC from 2007-06:2017-06
- Computed the median price for properties in each zip code of NYC for each year from June 2007 till June 2017
- Assigned 'Year' as the column name to the yearly median price per zipcode in NYC dataframe
- Concatenated the yearly median prices for each year from 2007-2017 with their respective zip codes.
- Restructured the dataframe to include RegionName, Year and MedianValue in the final Zillow dataframe as shown in the snapshot of first five rows below:

	RegionName	Year	MedianValue
0	10013	2007	2269050.0
1	10014	2007	1694500.0
2	10011	2007	1622800.0
3	10003	2007	1486700.0
4	10023	2007	1531350.0

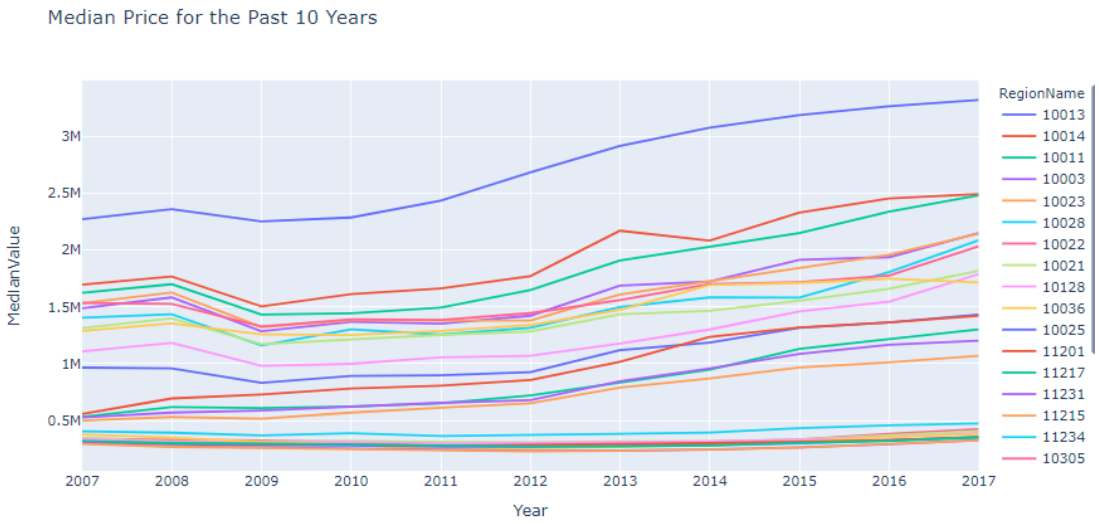
Zillow data exploratory analysis

First, I did the exploratory analysis on Zillow dataset.

Our initial Zillow dataset consisted of monthly median price which would have been cumbersome in performing further analysis on the dataset. So, I computed yearly median price for every zip code in NYC for the past 10 years. I selected median over mean to arrive at the average cost price in each zip code of NYC for the following two reasons:

1. Mean can easily be affected by outlier values (low or high monthly prices) in the dataset thereby skewing the Yearly price per zip code results.
2. During the period 2007-2010, USA was going through the subprime mortgage crisis that negatively affected house prices. In this case, the house prices will be on the lower side and taking a mean would not eliminate this bias.

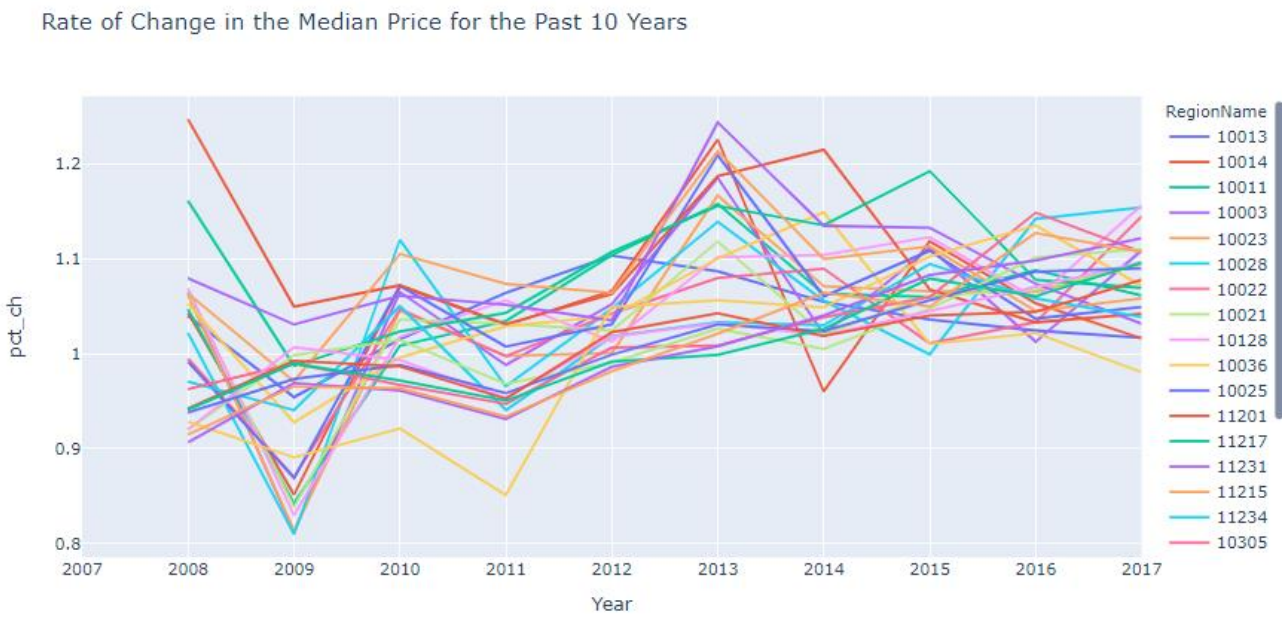
I begin my analysis to find out the trend in the median price (cost price) of each zip code in NYC for the past 10 years.



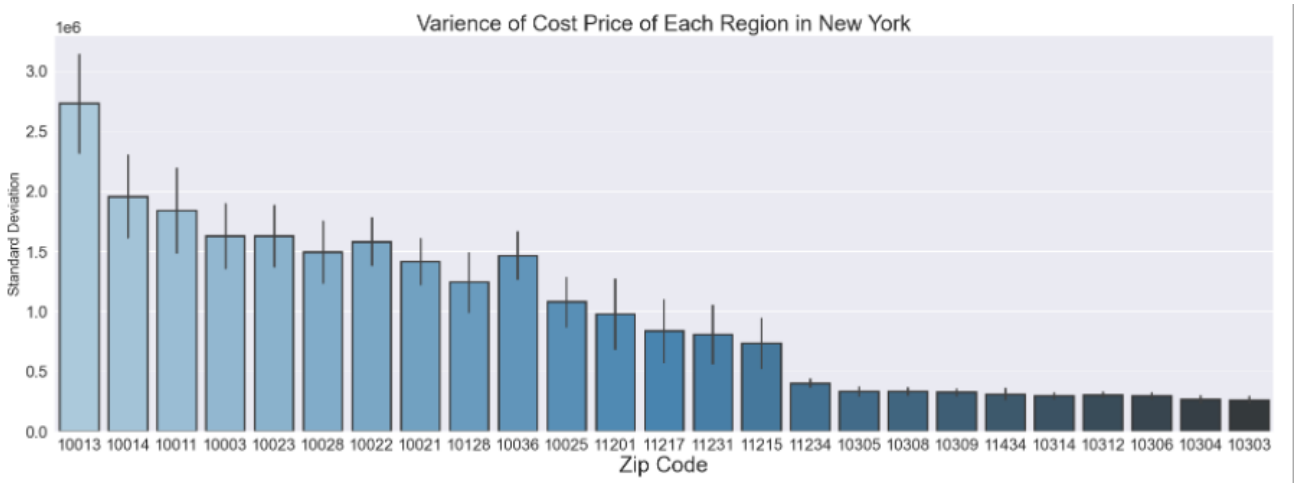
The Above trend lines of median price of each region name shows:

- The median price for the past 10 years is flat for some of the zip codes like 10303,10304,10305, 10306 etc.
- The median price for the past 10 years is volatile for some of the zip codes like 10013,10014,10011 etc.
- The median price for the past 10 years also shows that regions like 10013,10014,10011 have been the top 3 costliest regions for the past 10 years.
- The median price for the past 10 years also shows that regions like 10303,10304,10305, 103061 have been on the lower end in terms of cost for the past 10 years.

Then I tried to find out the trendline of the percent change in median house price per zip code in NYC.



From the above line chart, it is difficult to discern the volatility of the median price. Hence, I will next attempt to calculate the standard deviation in the yearly median house price per zip code.



The above bar graph certainly clarifies as to which regions are highly volatile and which ones are least volatile. Based on the standard deviation of the last 10 years median house price of the properties in every zip code of NYC, we can determine the following:

- Zip codes: 10013, 10014, 10011 are among the top three regions that have high volatility in their property price.
- Zip codes: 11234, 10305, 10308, 10309, 11434, 10314, 10312, 10303, 10304, 10306 have the least volatile median house prices over the last 10 years.

- Also, the regions with the highest volatility are also important to note since it increases the risk of investing in these zip codes since the prices can fluctuate drastically with time.

To predict the median price for the year 2018, 2019 & 2020, we calculated the average percentage change in the median price for the past three years (2015, 2016 and 2017) to maintain the recency of how median house prices have changed in each zip code. Once the 3-year average of the percentage change in median house price has been determined, we multiply the median cost price of 2017 with the 3-year average of the percentage change in median price to get the cost price for 2018. Similarly, we calculate 2019 prices by utilizing the calculated 2018 prices and finally utilize the 2019 prices to determine the 2020 cost price in each zip code.

Prediction of the median price for the year 2020 is needed to check the current median house price in NYC and evaluate the following:

- which regions are costly?
- which regions are comparatively cheaper in New York City?

The above bar graph is clearly showing that regions like 10013, 10014, 10011 are the costliest based on median price for the year 2020. Median price of 2020 is the current cost price for further analysis. Hence, we evaluated the zip codes in NYC based on our first key metric: Cost.

Airbnb dataset

Data cleaning and Processing of the Airbnb dataset

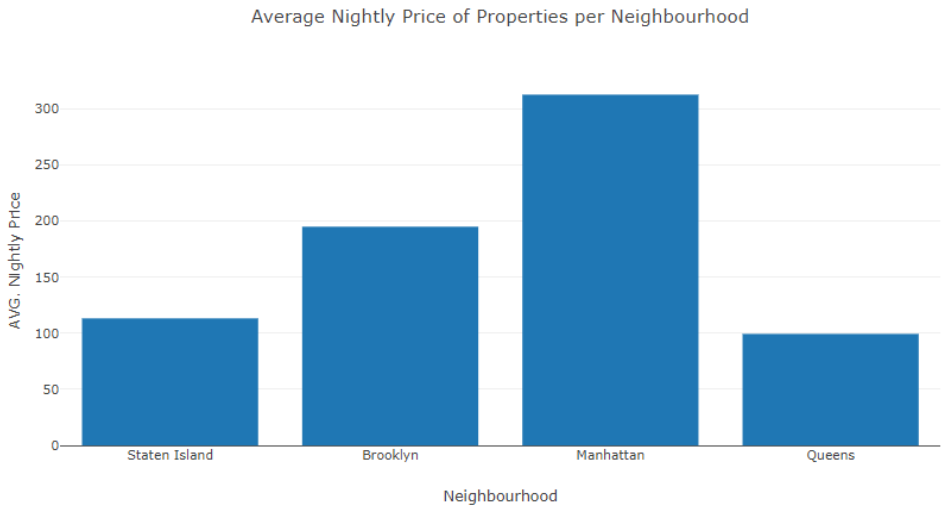
Following steps were taken to clean the Airbnb dataset:

- Changed the datatype of column 'id' from integer to string
- Removed the $ sign, empty spaces ,and "," from price column
- Changes the datatype of price from string to float
- Called missing_column_values function to compute the sum and percent of missing values of each column in the dataset
- Dropped all the columns with null value(s)
- Called missing_column_values function again to check if there are any columns with missing values.
- Merged the two dataframe based on inner join.
- Renamed the column MedianValue to Cost.

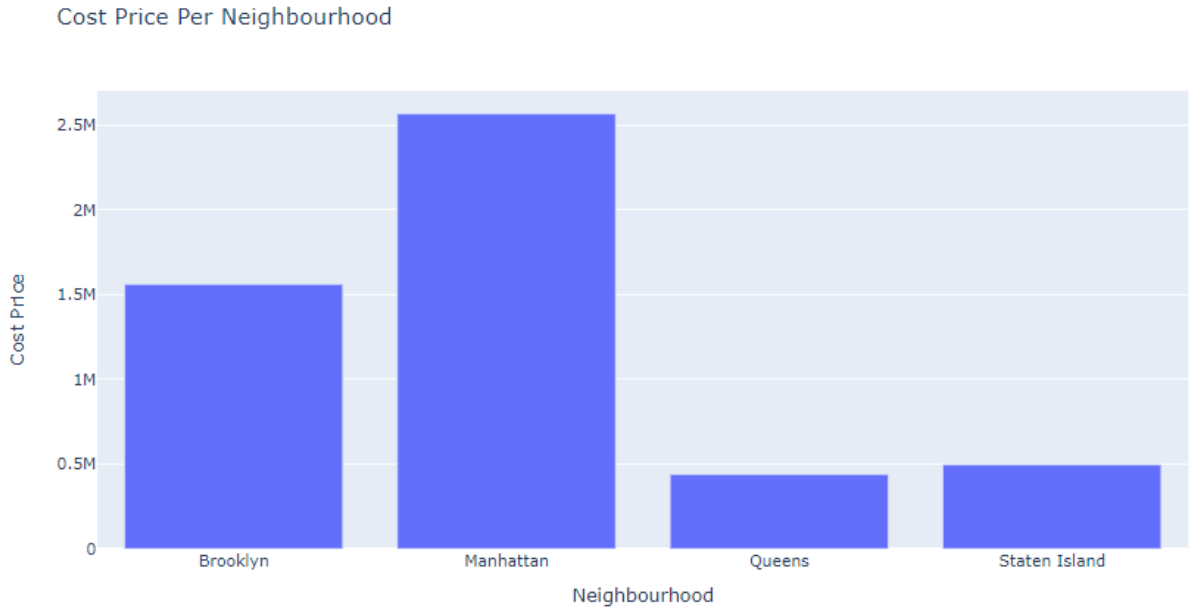
Exploratory analysis of the Airbnb dataset

After running exploratory analysis on Zillow dataset, we now want to do exploratory analysis of the Airbnb dataset. To perform this analysis, we first join the 2020 cost price per zip code in NYC data that was calculated using the NYC Zillow dataset with the Airbnb dataset.

We begin the analysis by finding out Average Nightly Price of Properties per Neighborhood.

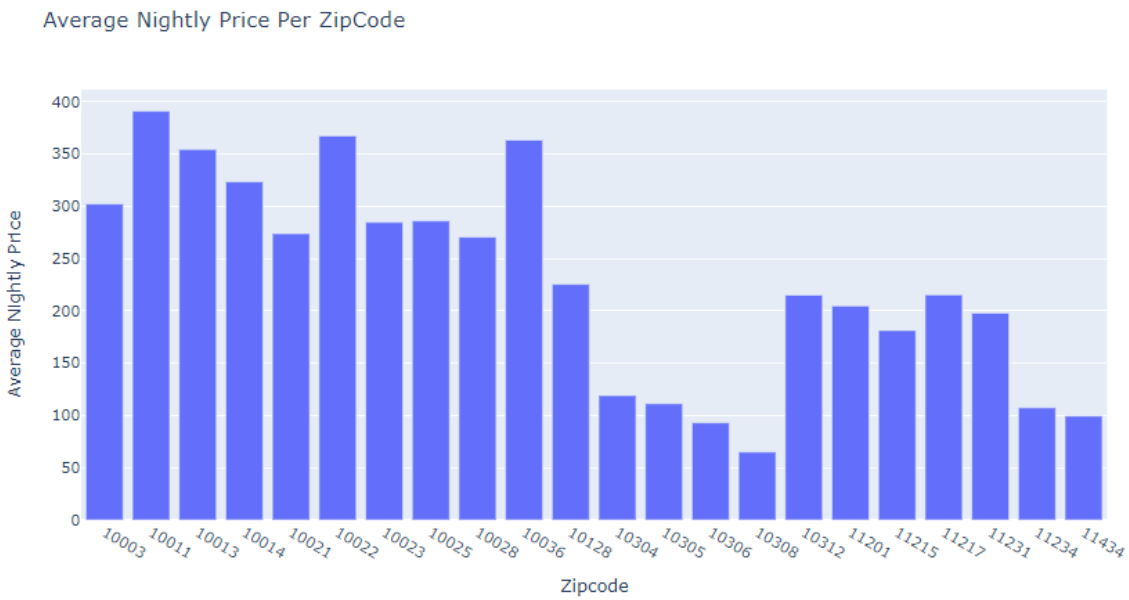


The above bar graph shows that the average nightly price is the highest in Manhattan neighborhood followed by Brooklyn, Staten Island, and Queens.
Then, we looked at the cost price in each neighborhood.



As per the cost price, Manhattan neighborhood is leading and followed by Brooklyn, Staten Island, and Queens. Although, Manhattan is the costliest neighborhood but revenue from rentals is also the highest. However, to arrive at which neighborhood is the most profitable one, we need to dig deeper to find out annual rate of return per neighborhood. The rental price in a neighborhood is strongly corelated to the cost of that property in the neighborhood and the cost price drives the rental rates as well in each zip code of NYC.

Next, we calculated the Average Nightly Price Per Zip Code.

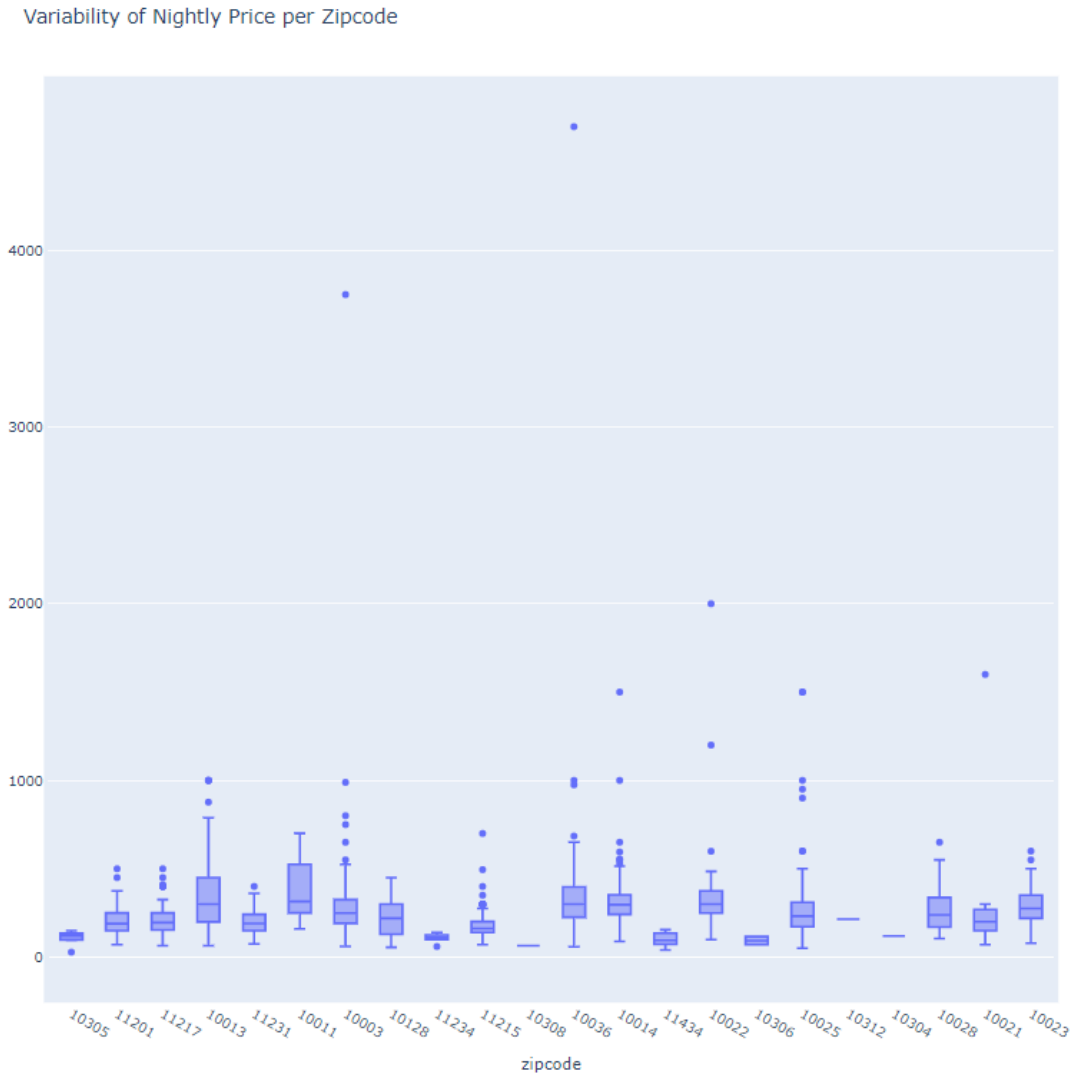


The above result shows which zip codes have the most and the least potential of earning daily rental revenue. Thereafter, tried to find out Average Nightly Price per Zip code by Neighborhood.

The above bar chart shows that out of 22 zip codes in the New York City dataset, Manhattan neighborhood has the greatest number of zip code and Queens has only one zip code. This clearly means that within Manhattan, there are a lot of popular zip codes where rental properties are in high demand. However, in Queen there is only a single zip code which is popular for rentals. For both Staten Island and Brooklyn, there are 5 zip codes each that are

popular for rentals. The above figure is also clearly showing as to which zip codes by neighborhood has the most potential for rental income and has the highest nightly rental prices. For example, zip code 10011 in Manhattan seems to be the most expensive in terms of average nightly rental price. However, to determine profits, we will also need to consider the cost price of the properties in each zip code and then compare it with the average nightly rental price.

Thereafter, we tried to find out the variability of nightly price per zip code as well as variability in nightly price per Zip code by neighborhood.

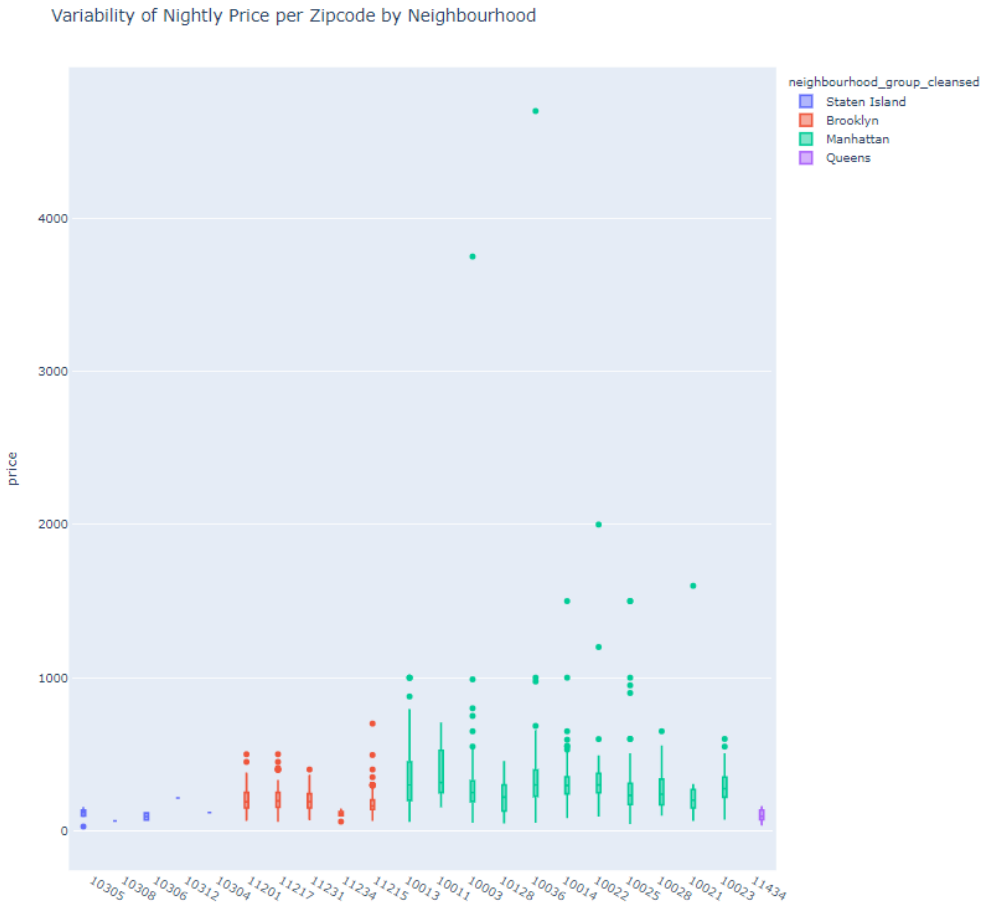


The above boxplots shows that the zip code 10036 has the highest nightly price range with min: 58 and max: 4700. The next zip code with the highest nightly price range is 10003 (min:

60, max: 3750). However, both zip codes have outliers which is causing the range to be inflated.

However, if we exclude zip codes with outliers, the zip codes with the highest variability are the following zip codes and I am listing them in descending order of variability: 10011, 10013, 10036 and 10128.

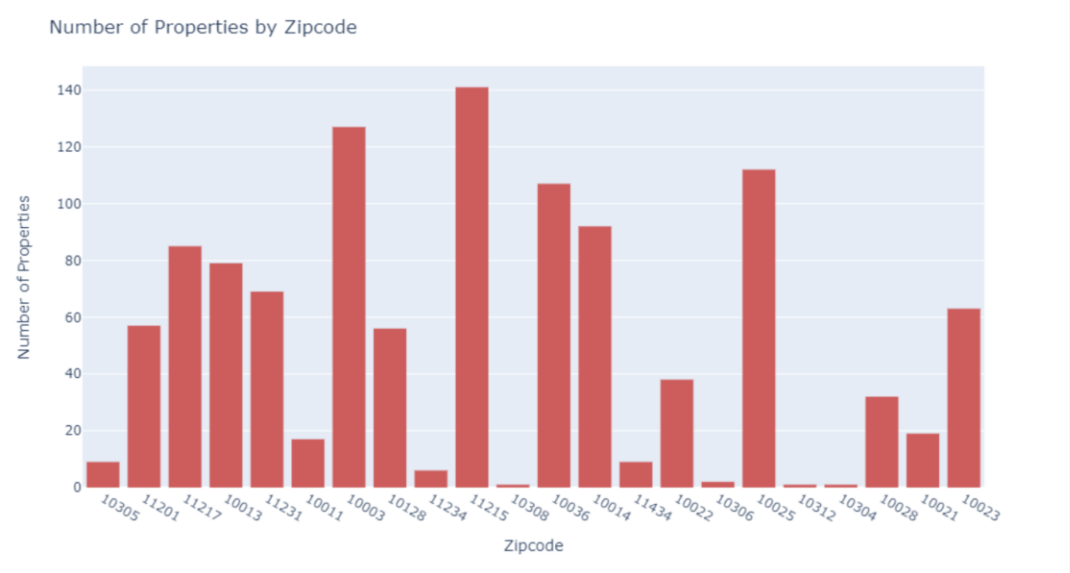
Zip codes such as 10312, 10304, 10308 do not have any price variability. We need to dig further to find out in which neighborhood they are falling.



The above results made it clear that zip codes falling in Manhattan neighborhood has the most price variability which indicates that the highest fluctuation in rental price is in that neighborhood. Neighborhood such as 'Staten Island' has the least variability of price. High nightly price variability also indicates that potential for higher risk in Manhattan due to fluctuations in price. However, the fluctuation in price might also indicate seasonality with

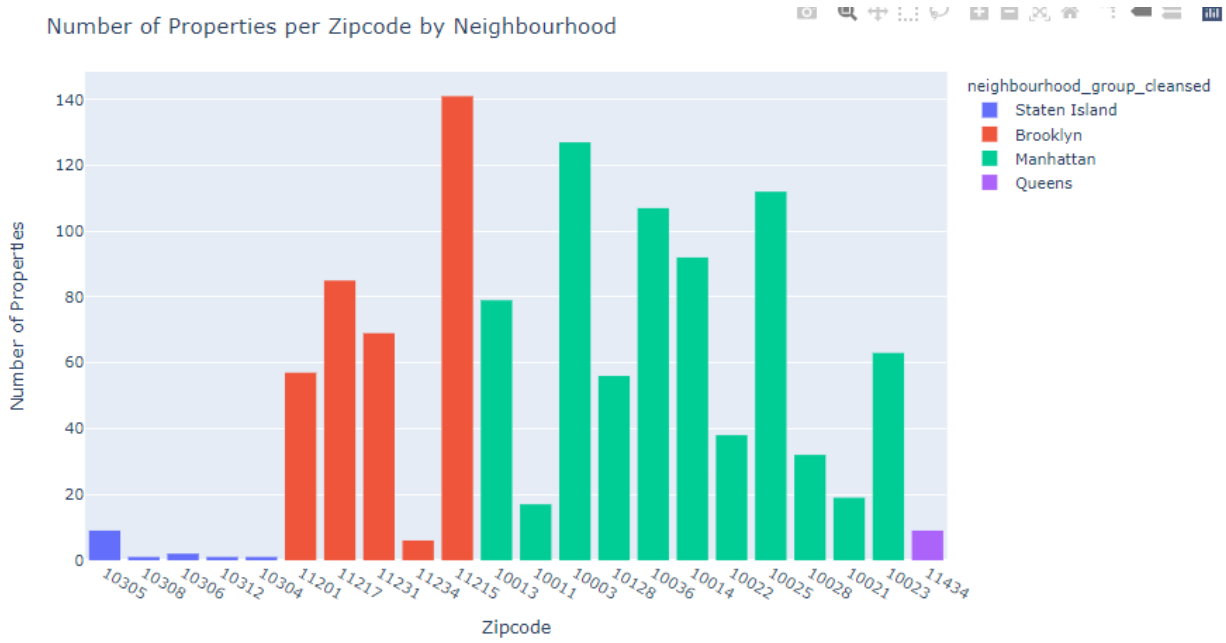
rental rates being higher in the summer as compared to winters due to variability in demand. Hence, we should keep in mind that the cash flow from rental properties in these two neighborhoods will fluctuate and should be factored into cash flow projections in the future.

Thereafter, we did analysis to find out number of properties by zip code as well as number of properties by zip code by neighborhood.



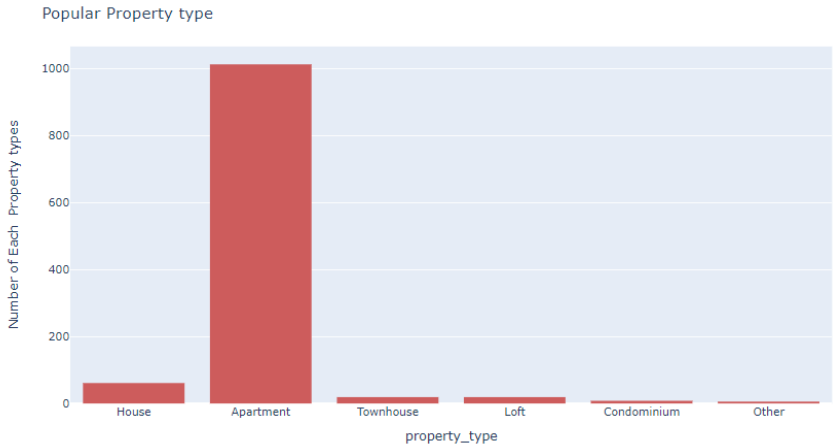
The above bar graph shows that the zipcode:11215 has the highest number of properties and zipcode:10312, 10304 has the least number of properties. We need to find out in which neighborhood these zip codes are located. But it is clear from this data that these are the most

popular zip codes for rentals.

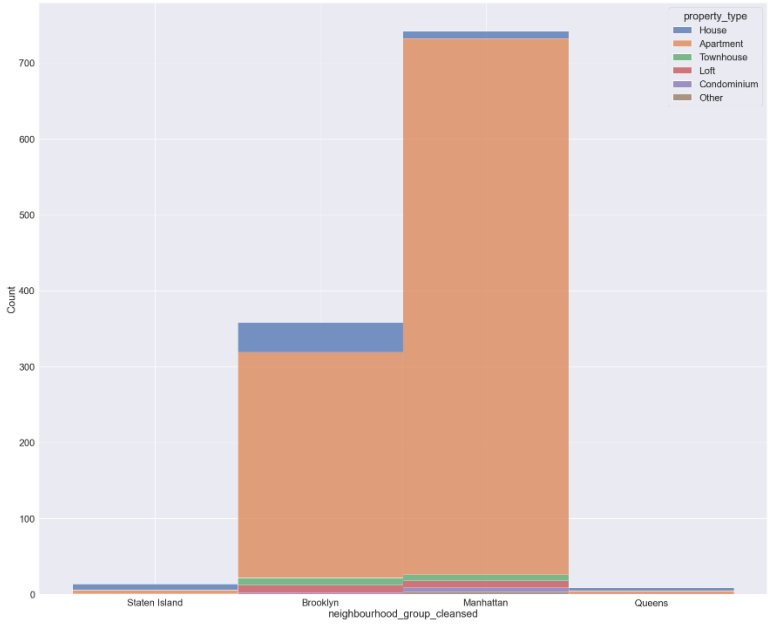


The above bar graph shows that the zip code 11215 which has the highest number of properties, is in Brooklyn neighborhood. The graph makes it clear that the neighborhoods that are the most popular in terms of short-term rentals are Brooklyn and Manhattan. The highest potential for investment in properties for short term rentals is in these 2 neighborhoods because they have the highest demand. We will need to check for the rate of return in these neighborhoods by zip code to determine the most profitable zip codes to invest in.

Next, we investigate the most popular property type for short-term rental in NYC.



The above result indicate that apartment is the most popular property type by a long distance. We will next determine which neighborhoods have the most apartments.



It is clear from the above graph that Manhattan has the highest number of apartments followed by Brooklyn. Hence, Brooklyn and Manhattan are the most popular neighborhoods with the greatest number of apartments available for short-term rentals (which is the most popular property type for short term rentals). Hence, it is prudent to invest in apartments type properties in Manhattan Brooklyn. It also indicates that Manhattan and Brooklyn should have enough apartment type properties available, and it should be the highest priority to invest in those properties.

Key findings

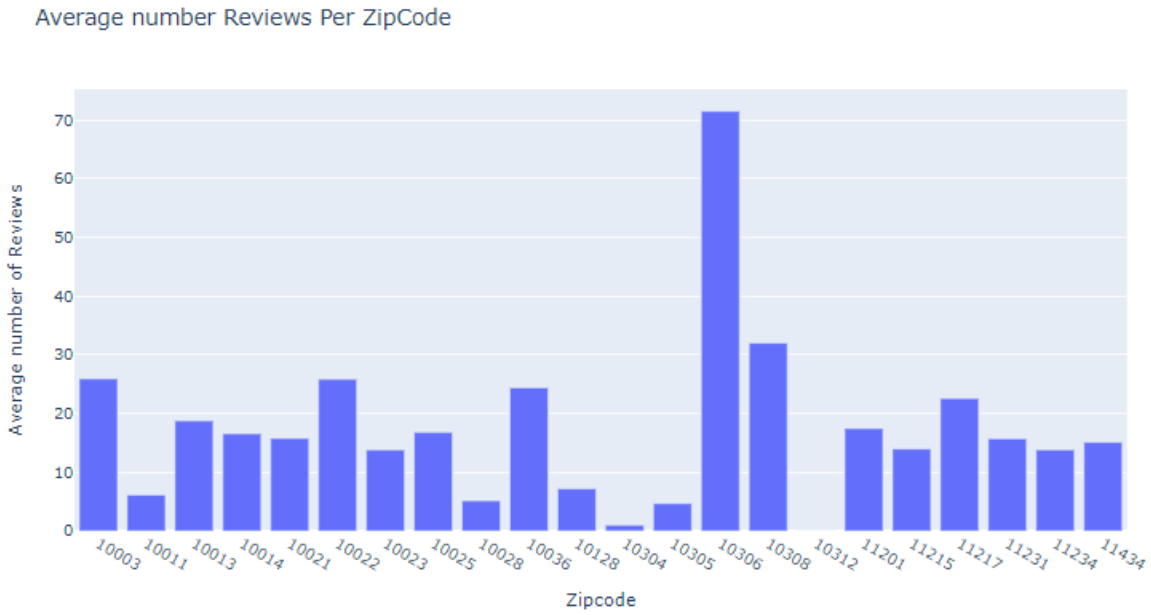
- Manhattan and Brooklyn are the most popular neighborhoods for rentals. Within Manhattan, there are multiple zip codes that are popular, but Brooklyn only has five zip code that are popular for short term rentals.
- Looking at the count of rentals also confirms that Manhattan and Brooklyn are by far the most popular for short-term rentals.
- Apartments is the most popular property type and Manhattan is the neighborhood that has the highest number of apartments followed by Brooklyn.

- The variability is rental prices in Manhattan and Brooklyn might be a concern since it might add risk for the investment since rental prices fluctuate widely. However, if the fluctuation in rental prices is due to seasonality, then it is something to factor since it will mean that the cash flow will vary by seasons. Hence, we should keep in mind that the cash flow from rental properties in these two neighborhoods will fluctuate and should be factored into cash flow projections in the future.

Comparative Analysis based on Key Metrics.

Finally, we tried to evaluate the zip code based on following key metrics:

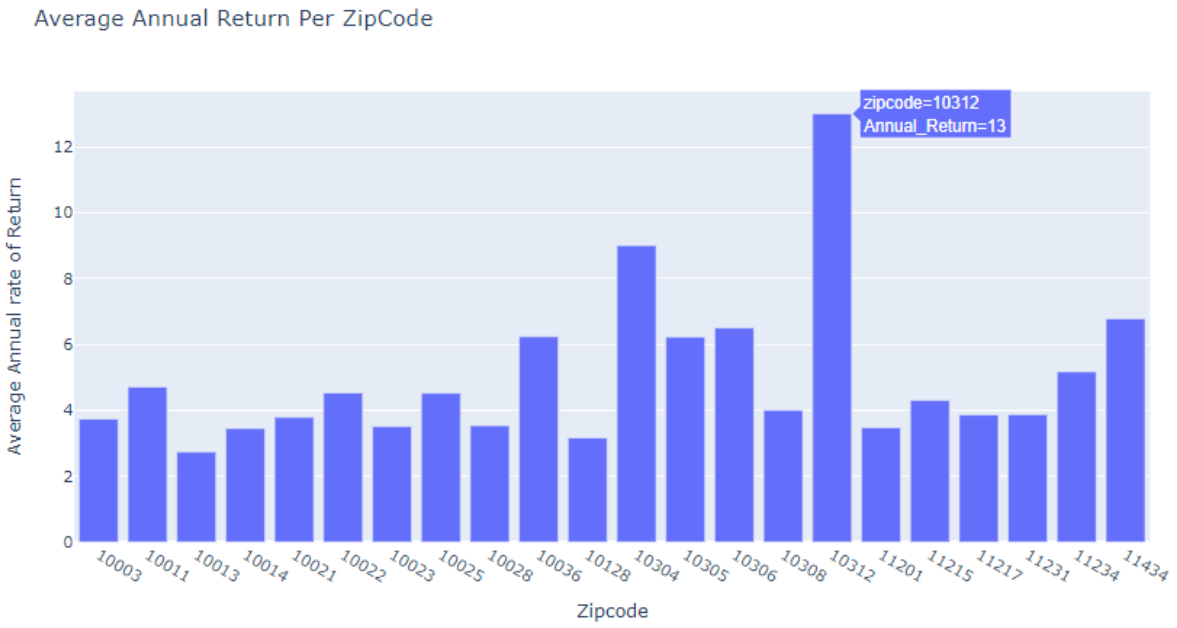
- Popularity by zip code: We evaluated zip codes based on average number of reviews per zip code. The higher the average of the number of reviews, higher is the demand for that zip code. The reviews provide information about which properties are favorable among renters.



The above result shows the zip codes that are in top five in terms of popularity are: 10306, 10308, 10003, 10022,10036.

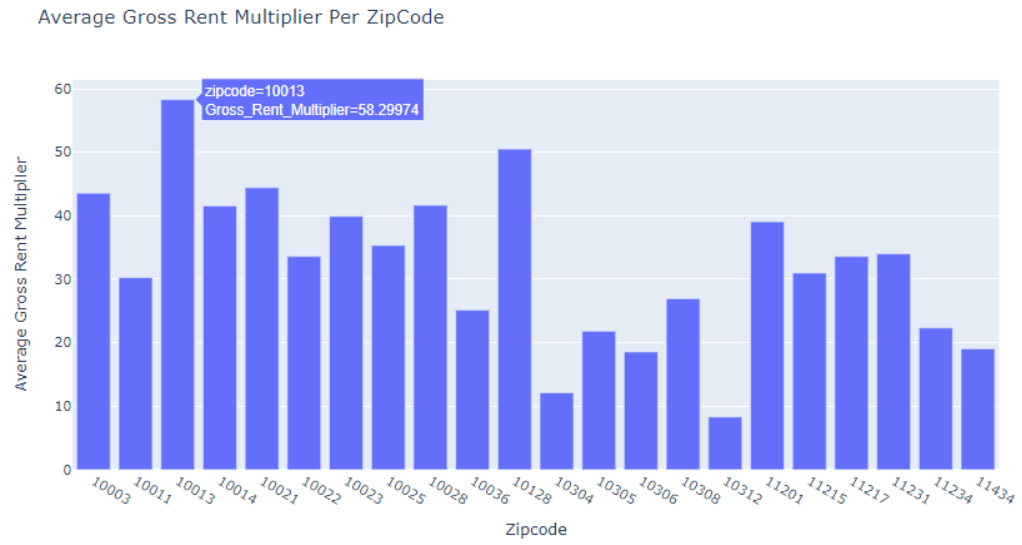
- Annual Rate of Return per zip code

  To calculate the annual rate of return, we first calculated the annual rent by multiplying nightly price with occupancy rate (75% of 365 days a year). Then calculated the percentage: (annual rent/cost) * 100.



As per the above results, zip code 10312 has the highest average annual rate of return with almost 14% followed by the zip code: 10304(9%), 11434(7%), 10036(6.2%).

- Gross Rent Multiplier



Gross Rent Multiplier is calculated by using following formula:

(Property Price / Gross Annual Rental Income)

The Gross Rent Multiplier calculation compares the property's cost price or fair market value to the gross rental income. Using the gross rent multiplier is a good way to take a "quick look" at how fast the property will be paid off from the gross rent the property is generating. Lower the better. Based on this metric, zip code 10312 is leading. Instead of using this metric, we will leverage the Annual rate of return metric calculated above.

Conclusion

To evaluate in which zip codes to invest, we will measure how each zip code performs against each of the following metrics:

1. Cost: The cost of property in each zip code. The lower the cost, the better the zip code is in terms of cost.
2. Popularity: The average number of reviews each zip code received. The average number of reviews indicate that it is a popular zip code for rentals and the higher it is, the better. Also, it is a reflection of the favorability of the zip code among renters.
3. Annual rate of return: This key metric is considered to find out how much rate of annual return is to be expected from each zip code. The higher it is the better the zip code as per this metric.
4. Variance in rental price: This metric measures how much variability is in the rental price for each zip code. The lower this metric, lower is the risk of investment and more stable will be the cash flow and revenue from these zip codes.
5. Most options: This metric measures the zip codes with the greatest number of rental properties thereby representing the zip code with the most number of options to rent a property for a short duration.

The following table is showing the top 10 zip codes for each of the above 5 metrics.

Cost	Popularity	Annual Rate of Return (Annual ROR)	Variability in rental prices	Most options
10304	10306	10312	10308	11215
11434	10308	10304	10306	10003
10303	10003	11434	11434	10025
10314	10022	10306	11234	10036
10306	10036	10036	10305	10014
10308	11217	10305	10304	11217

10309	10013	11234	11215	10013
10312	11201	10011	11231	11231
10305	10025	10022	11201	10023
11234	10014	10025	11217	11201

Among the 10 zip codes for each category, the next step is to determine which zip code satisfies at least 3 of the metrics and idea is to prioritize the investment in those zip code. Based on this analysis, the following zip codes at least satisfy 3 of the above metrics.

Neighborhood	Zip codes	Number of metrics
Staten Island	10304	3
Queens	11434	3
Staten Island	10306	3
Staten Island	10308	3
Staten Island	10305	3
Brooklyn	11234	3
Manhattan	10036	3
Brooklyn	11217	3
Brooklyn	11201	3
Manhattan	10025	3

In conclusion, my methodology resulted in zip codes that satisfy 3 of the metrics at most. Based on this analysis, the real estate company should prioritize investment in the following zip codes in NYC for short-term properties:

1. Zip code 10304,10306, 10308 and 10305 in Staten Island. The zip codes in Staten Island qualified because they performed very well in the “Cost”, “Annual ROR” and “Variability in rental prices” metrics.
2. Zip code 11434 in Queens also performed well in the “Cost”, “Annual ROR” and “Variability in rental prices” metrics and hence qualified as a good zip code to invest in.
3. Zip code 11234, 11217 and 11201 in Brooklyn. Zip code 11217 and 11201 qualified since they did well in “Popularity”, “Variability in rental prices” and “Most Options”. Zip code 11234 on the other hand did well in “Cost”, “Annual ROR” and “Variability in rental prices”.
4. Zip code 10025 in Manhattan. The only zip code form Manhattan qualified based on performing well on the following 3 metrics: “Popularity”, “Annual ROR” and “Most Options”.

Further Improvements

- To predict median price or the cost price of each zip code, time series analysis could have been used instead of average percentage change in price for the past three years.
- Variables like longitude and latitude could have been used to show zip codes through geographical maps.