

# New York area analysis

- A Data Science Report –

## I. Introduction

Client A is a representative of a multi-national hotel chain group and he has been sent to a Data Science specialist in order to provide a report about the current situation of the hotels in the New York area. He requires information about venues like restaurants and other hotels. He is especially interested in the localization of a future hotel and is asking the Data Scientist for a suggestion based on an analysis of the venues.

He is looking forward to finding a good place that would attract customers for a new hotel that the company wants to build and he wants the report to be based on statistical data about the New York area.

## II. Business Problem

In order to provide recommendations of ideal places for building a hotel, I will use Foursquare location data. More specifically, I will import data about the venues (first of all the hotels from the New York area, and then other type of venues like restaurants, parks, etc.) and then I will concatenate the results with other New York datasets like Crime & Arrest data. I will use these datasets in order to provide an accurate suggestion based on the popularity of the area (number of venues, localization in New York) and crimes committed in the city because the investment is considerably large and crime data has to be taken into account. These statistics should provide a good indicator for a future hotel.

## III. Data

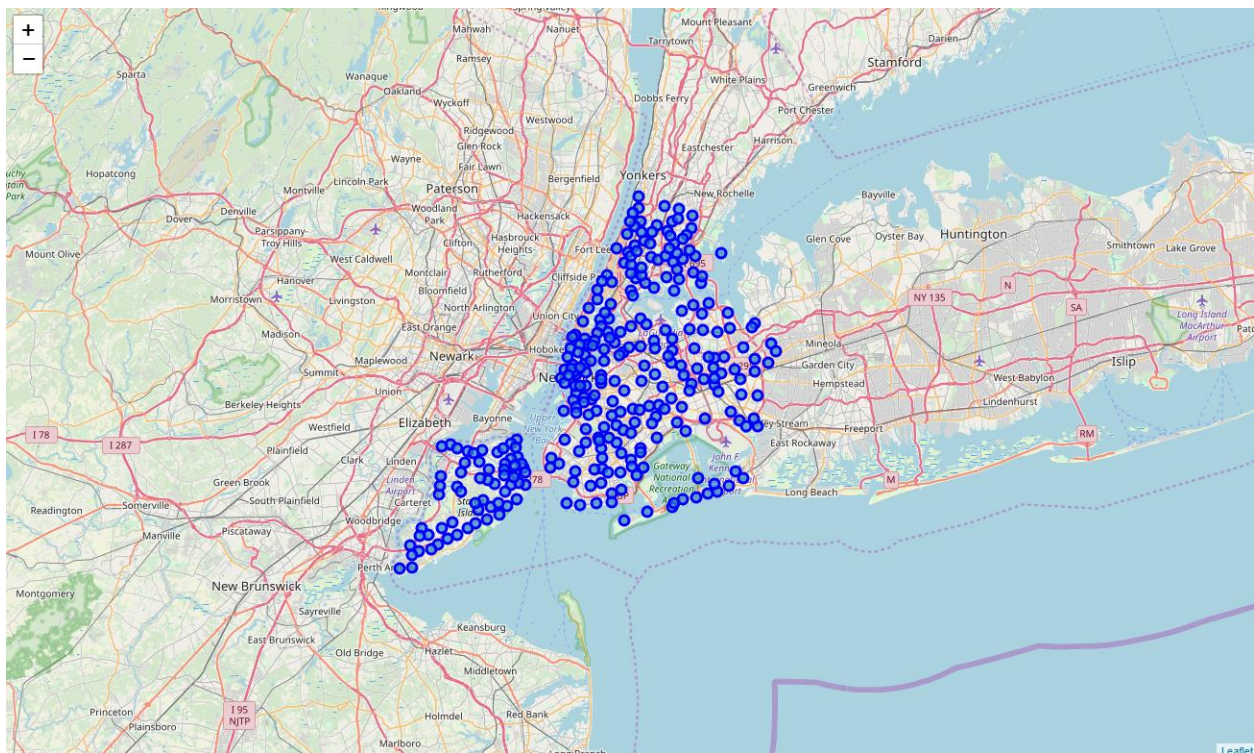
Data sources for this report are: Airbnb dataset, New York hotels dataset and NYPD crime complaints dataset from [www.kaggle.com](http://www.kaggle.com). Foursquare API has also been used in order to acquire information about venues based on coordinates. These sources will help me outline some locations where the company should build a hotel. I will analyze data about Airbnb properties locations, New York hotels locations and NYPD reported crimes in the New York area because all these aspects are relevant when researching for the optimum place to build a hotel.

## IV. Methodology

I have started by importing the data about the New York neighborhoods from a dataset. I kept only the columns that I would use for plotting a map.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

I was interested in finding a good place for a building a hotel and the map was useful when analyzing upsides and downsides of different areas.

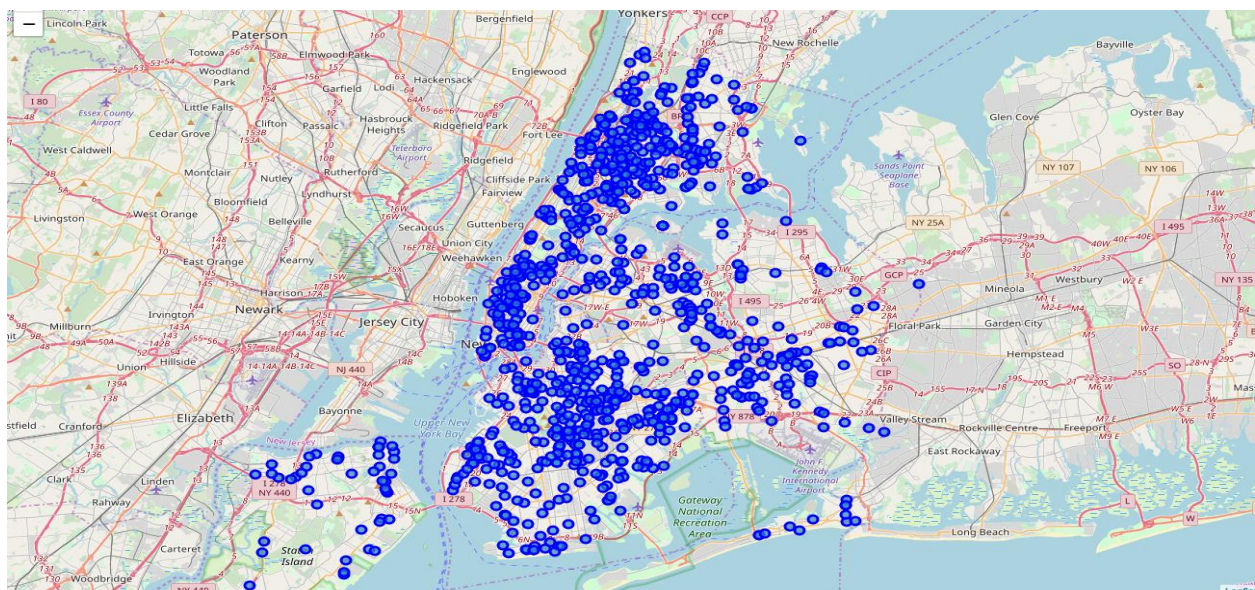


I have imported a dataset about crimes complaints from NYPD and analyzed the results. In order to be able to plot a map from the dataset, I had to create a new dataframe containing coordinates. If we import data from NYPD crimes complaints we can observe that there are several

neighborhoods in which the crime rate is higher. Judging by 1000 crimes committed in New York City we can assume that certain areas from Staten Island are safer than the rest. Neighborhoods along the Central Park are also in the same situation. The number of crimes from the other zones in New York is bigger than in those areas.

	Description	Latitude	Longitude
0	FORGERY	40.828848	-73.916661
1	MURDER & NON-NEGL MANSLAUGHTER	40.697338	-73.784557
2	DANGEROUS DRUGS	40.802607	-73.945052
3	ASSAULT 3 & RELATED OFFENSES	40.654549	-73.726339
4	ASSAULT 3 & RELATED OFFENSES	40.738002	-73.987891
...	...	...	...
1010	OFF. AGNST PUB ORD SENSBLTY &	40.808374	-73.926996
1011	FRAUDS	40.863773	-73.922196
1012	OFF. AGNST PUB ORD SENSBLTY &	40.701986	-73.927964
1013	OFF. AGNST PUB ORD SENSBLTY &	40.718564	-73.988201
1014	CRIMINAL MISCHIEF & RELATED OF	40.688201	-73.988084

I turned this dataset into a map that is displaying the crimes committed in New York.





We can already observe that there are some neighborhoods in which a hotel would be safe.

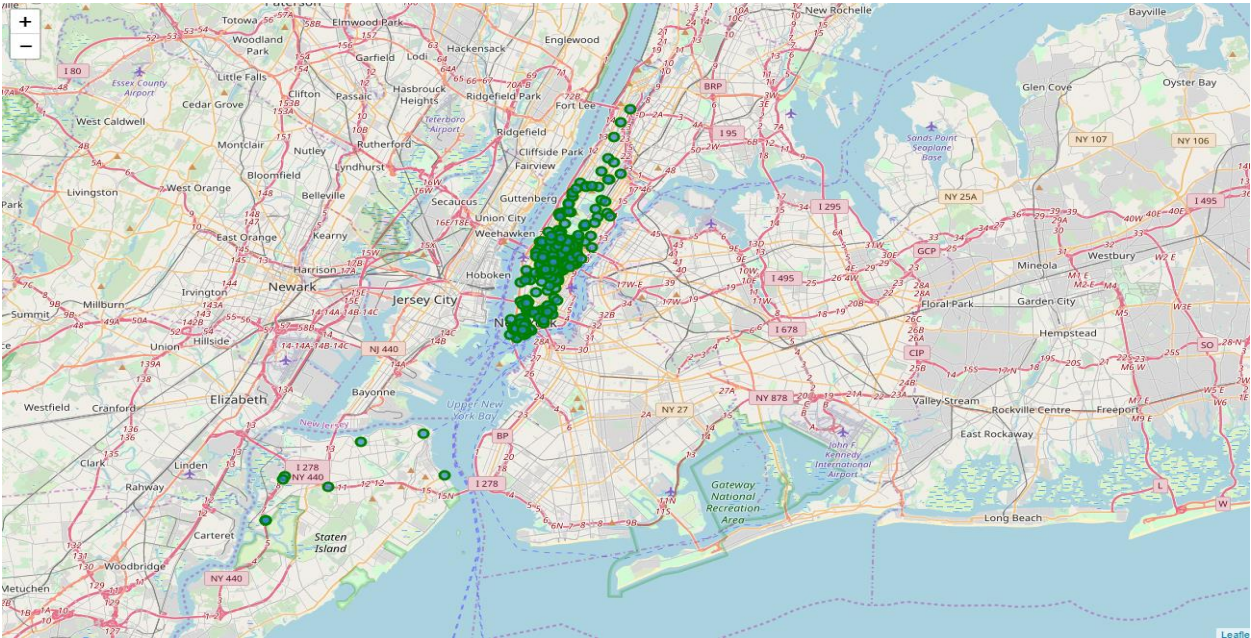
Another important aspect when choosing a place for the hotel is the competition of that area.

Therefore I created a dataframe with hotels and Airbnb listed properties.

city	latitude	longitude	neighbourhood	latitude	longitude	availability_365
New York	40.74624	-73.97929	Kensington	40.64749	-73.97237	365
New York	40.76452	-73.98078	Midtown	40.75362	-73.98377	355
New York	40.76366	-73.97965	Harlem	40.80902	-73.94190	365
New York	40.74929	-73.97670	Clinton Hill	40.68514	-73.95976	194
New York	40.72102	-74.00418	East Harlem	40.79851	-73.94399	0
...	...	...	Murray Hill	40.74767	-73.97500	129
Staten Island	40.61286	-74.17925	Bedford-Stuyvesant	40.68688	-73.95596	0
Staten Island	40.64200	-74.07902	Hell's Kitchen	40.76489	-73.98493	220
Staten Island	40.61548	-74.06338	Upper West Side	40.80178	-73.96723	0
Staten Island	40.63676	-74.12343	Chinatown	40.71344	-73.99037	188
Staten Island	40.58683	-74.19143				

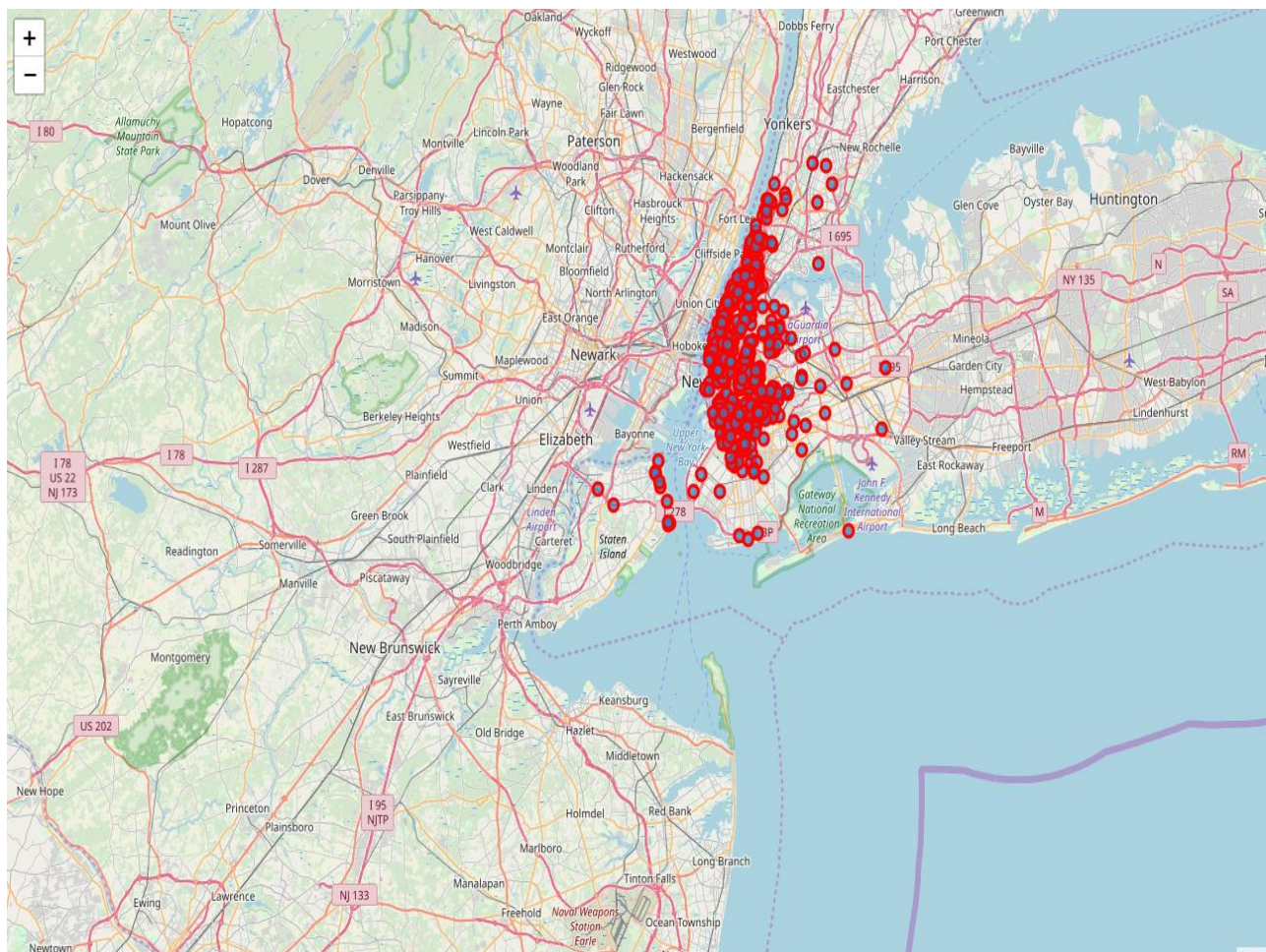
These dataframes resulted in a map that shows the placement of each hotel and Airbnb property.

According to the dataset, there are over 1600 hotels in the state of New York, with over 400 of them being in the New York Area (New York City + Staten Island).



However, when the map of hotels was created it showed a big discrepancy between the number of hotels in Staten Island and the number of hotels in New York City. New York City is crowded with hotels around the downtown area, which means that the investment will be higher if a hotel will be built in this area. In this case, however, the number of clients and most likely the revenue will increase drastically. Staten Island appears to be a more relaxed zone, with the hotels being more dispersed.

Hotels are not the only places tourists can rent a place. Nowadays a multitude of people choose other methods of finding a place to rent. Such an example would be Airbnb. One aspect that is worth mentioning about Airbnb is that not all properties are available every single day of the year. In fact, the properties listed on Airbnb are available, on average, 212 day a year. These are mainly concentrated in New York City, with few of them being in Staten Island. Companies that own hotels should be aware of the threat that Airbnb poses for their businesses, but there are many perks that hotels have over the Airbnb properties: usually all-year availability, room-service, breakfast and restaurants.

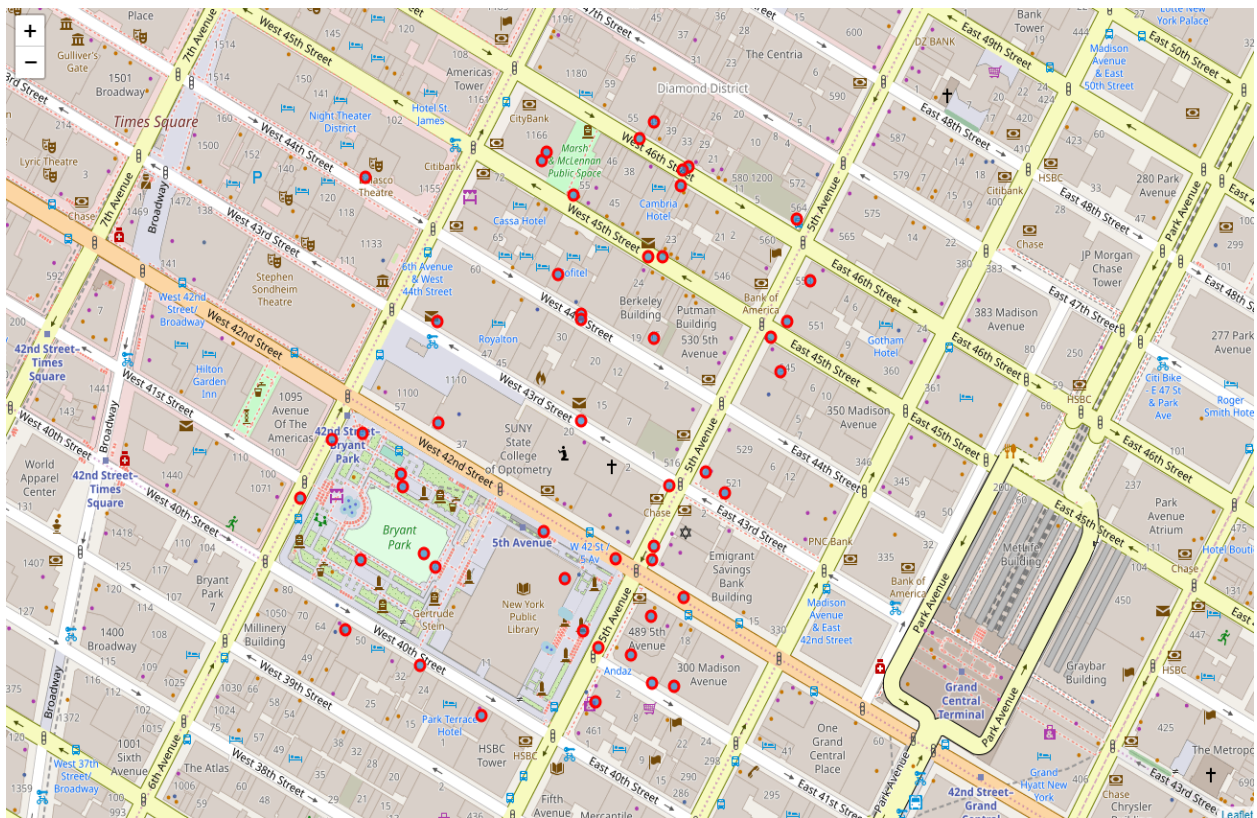




```
short_air1['availability_365'].mean()
```

212.03703703703704

New York is crowded with venues that are meant to attract you. For tourists, venues are points of interest when visiting because they want to spend some time relaxing. For this reason, it is important to have venues close to the hotel. Downtown New York provides probably the best experience for those wanting to go out, so the place is perfect if when building a hotel you're looking for client satisfaction. In a 500 meters radius there are plenty of venues to be found. The neighborhood displayed in the map is Midtown, and has the largest population density. It is an interest point for every tourist wanting to visit New York.



In order to plot this map, I have accessed Foursquare's database through its API and called for all venues in a 500 meters radius.

Staten Island lacks the diversity of venues that you can find in New York City, but it still provides a good number of different attractions. It is categorized as a more relaxed zone, attracting tourists through its simplicity. The neighborhood shown in the map is New Brighton, which is in the Northern part of Staten Island. It is the zone that is closest to New York City and the crimes rate here are low, the listed Airbnb properties are scarce and the competition from other hotel companies is not as powerful as in Midtown New York. We can find a number of venues in this neighborhood: parks, stores, restaurants.



## V. Results

While New York City provides a great number of venues and attractions for tourists, there is a price for that. Because of the high population density areas in New York we can pinpoint many crime complaints and competitors. On the bright side New York City, especially downtown, is an economic hub that attract people that are willing to spend money for the experience. While there are numerous downsides to owning a hotel in downtown, there is also a high reward that comes with it. Staten Island has proven to be safe, quiet, relaxing and very close to New York City. The downsides here are that the number of clients will not be as big as in Midtown for example, but the costs of building and maintaining a hotel will be considerably lower.

## VI. Discussion

There is no perfect place to build a hotel, especially in New York, but these datasets helped me understand different problems that the city portrays. This report helped me have a critical thinking of data in general and how we can use it.

## VII. Conclusion

Based on the results from these datasets, we can conclude by saying that one investment is riskier than the other. Building a hotel in New York City is a big commitment. It comes with perks, such as customer satisfaction and overall rating, but there are many competitors in the area. Unfortunately, because of the number of Airbnb properties and the number of hotels that would be nearby the investment is likely to have bad results.

My recommendation is to proceed with investing in a building in Staten Island. A zone that is safer than Midtown and that provides a good experience to tourists because it is very close to New York City. The downsides in this case are few and risk decreases significantly.