

Project 3

**Prediction of the number of
travelers by Airbnb in New York City**

Diane Deroualle – May 2020

Problem Statement

Since 2008, Airbnb has been changing the way we travel around the world by offering solutions to stay in homestay accommodations. As one of the most visited cities in the world, New York City has plenty of accommodations to book through Airbnb.



➔ The purpose of this project is to construct a model with machine learning to predict the number of travelers per Airbnb in New York City considering the neighborhood, the room type, the price, the number of reviews, the availability...

Dataset

➤ dataset from www.kaggle.com with information about Airbnb metrics in New York City in 2019.

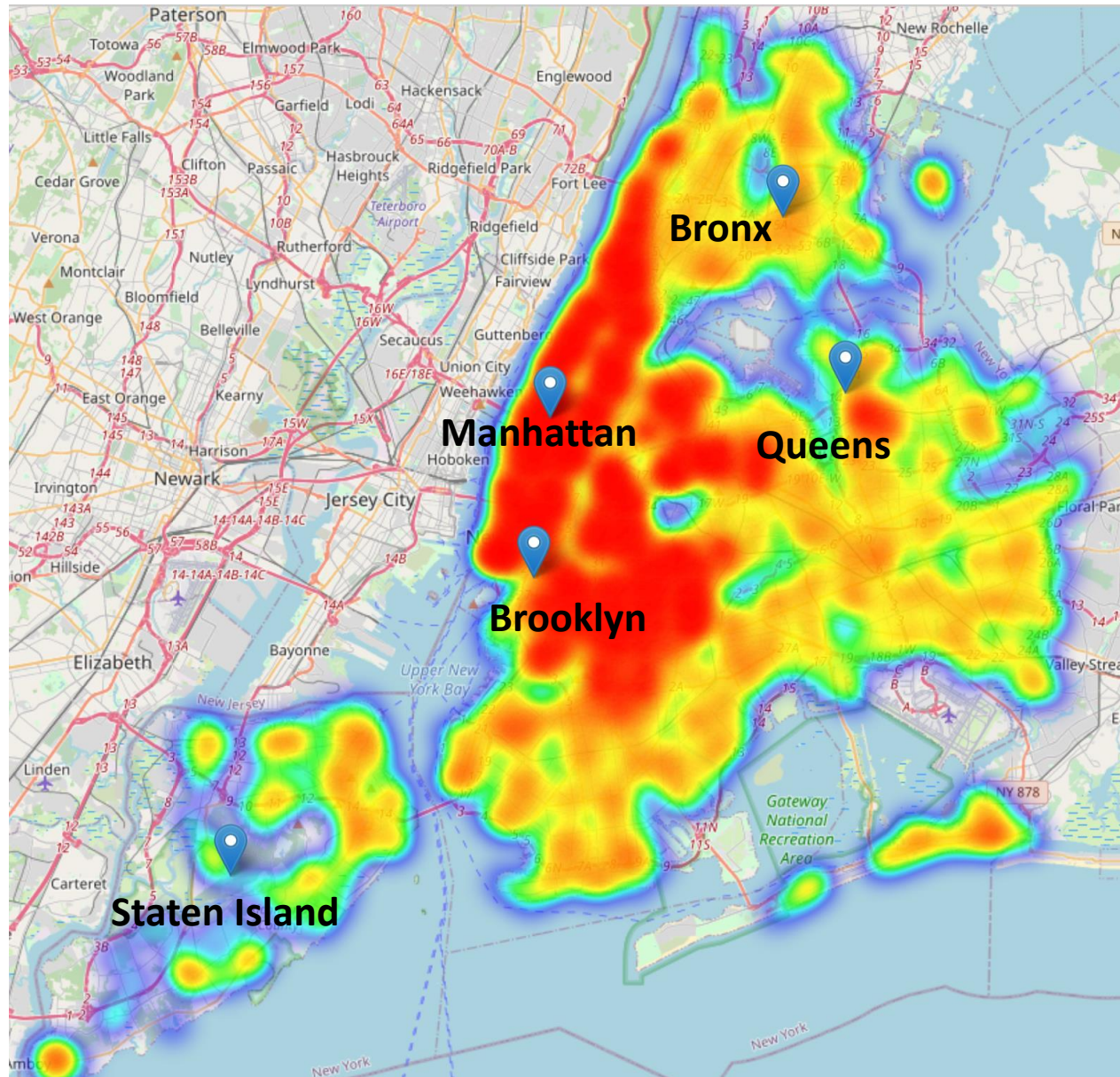
- Id
- Name
- Host_id
- Host_name
- **Neighborhood_group**
- Neighborhood
- **Latitude**
- **Longitude**
- **Room_type**
- Price
- **Minimum_nights**
- **Number_of_reviews**
- **Last_review**
- Reviews_per_month
- **Calculated_host_listings_count**
- **Availability_365**

➤ Dataset with 48,895 rows

Wrangling data

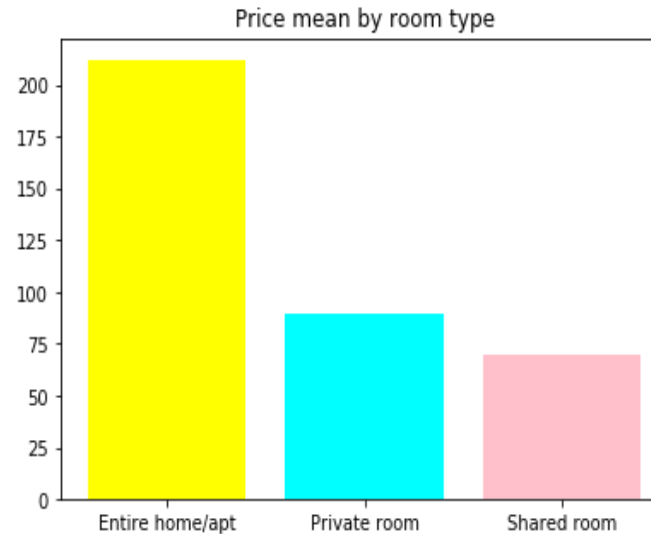
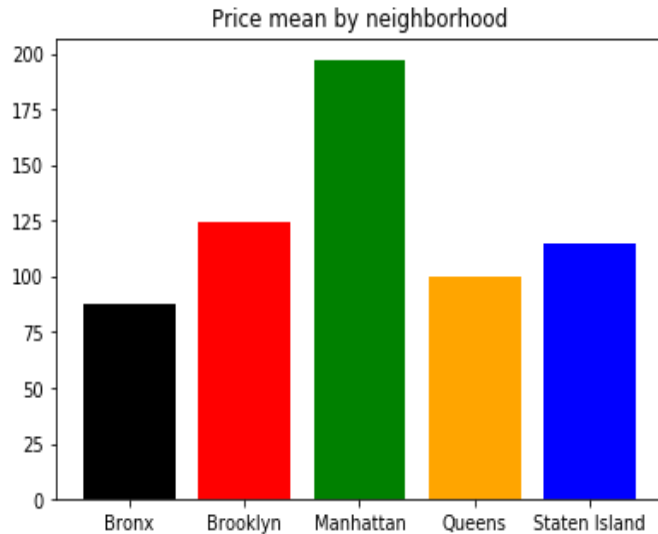
- Deleting unnecessary columns: *'id'*, *'name'*, *'host_id'*, *'host_name'*, *'neighborhood'* and also *'reviews_per_month'*
- Converting categorical features *'neighborhood_group'* and *'room_type'* into numerical columns
 - *'neighborhood_group'* = *'Bronx'*, *'Brooklyn'*, *'Manhattan'*, *'Queens'* columns (*'Staten Island'* deleted)
 - *'room_type'* = *'Entire home/apt'*, *'Private room'* (*'Shared room'* deleted)
- Deleting of rows with missing values in *'last_review'* column (10,052 rows)
- Converting *'last_review'* into the number of days since the last review: *'days_since_last_review'*
- Converting *'latitude'* and *'longitude'* columns into the distance with the center of Manhattan using the coordinates: *'latitude_from_Manhattan'* and *'longitude_from_Manhattan'*.

Heatmap number of travelers per Airbnb

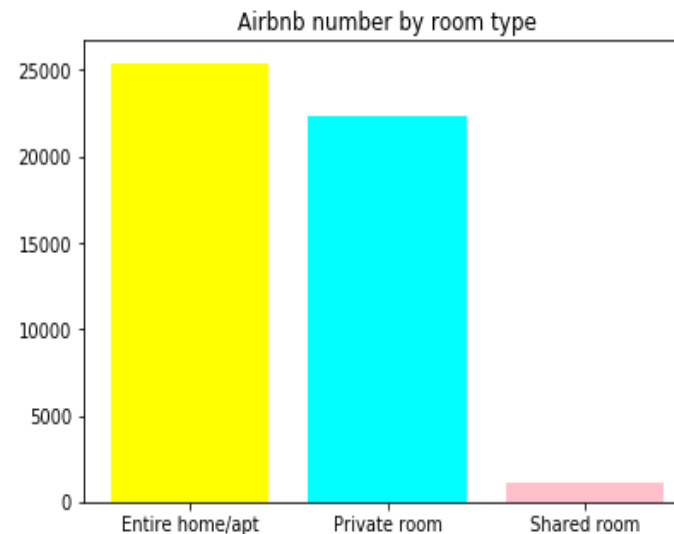
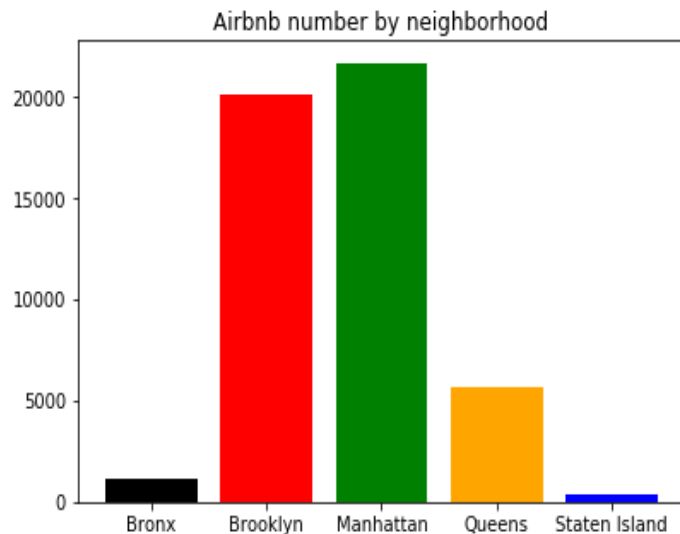


➔ Most of the travelers chose an Airbnb located in Manhattan, in the North of Brooklyn or the west of the Queens.

Visual exploratory data analysis

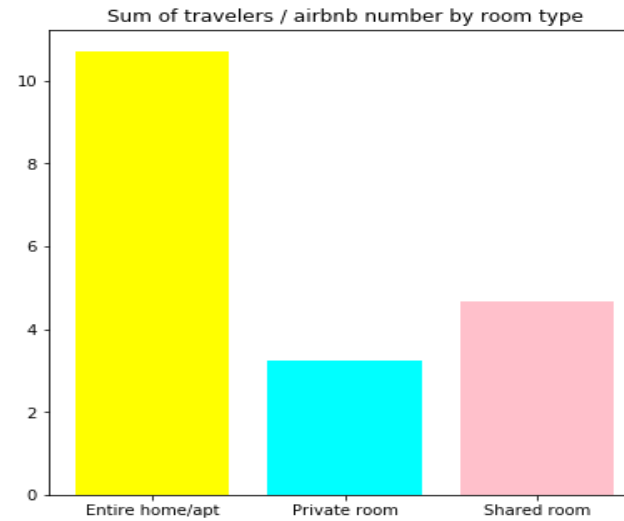
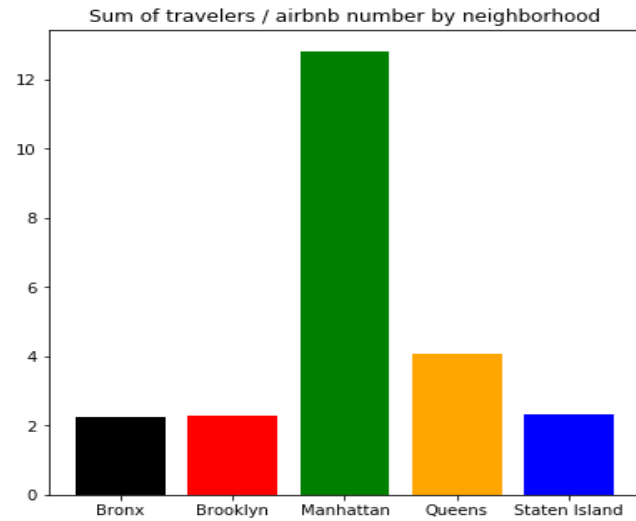


➔ Prices for the Airbnb entire houses/apartments on Manhattan were much higher than the others.

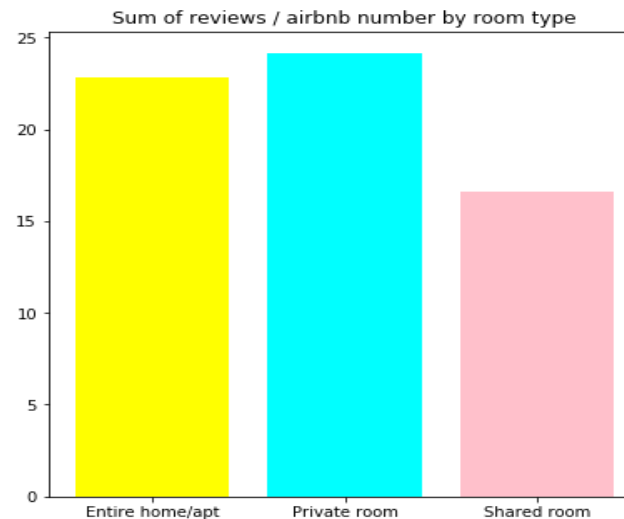
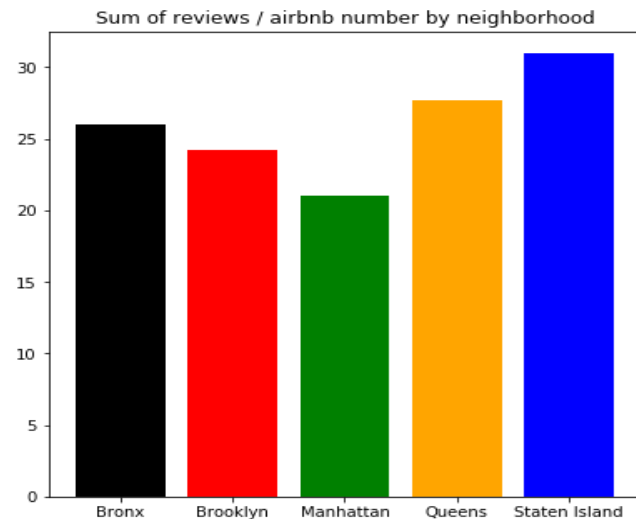


➔ There was much more Airbnb in Manhattan and in Brooklyn than in other neighborhoods. And almost all of them are entire home or apartment and private room.

Visual exploratory data analysis

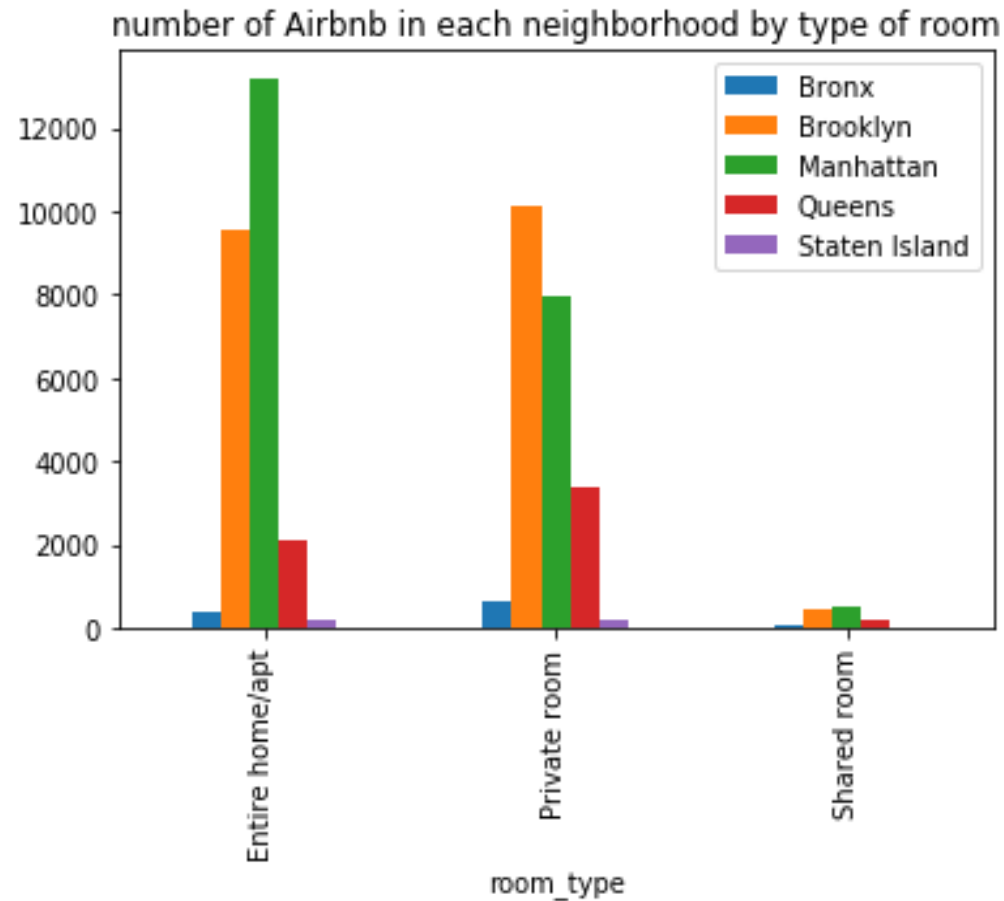


➔ Travelers preferred an Airbnb located in Manhattan and an entire home or apartment.



➔ Airbnb located in Manhattan had less reviews and Staten Island had more reviews. The entire homes or apartments and the private rooms had also more reviews.

Visual exploratory data analysis



➔ There were more entire homes or apartments than private rooms or shared rooms in Manhattan. In all other neighborhoods, this was the opposite.

Linear Regression Model

Features	+ / -	Coefficients
Minimum_nights	+	0.0540
Availability_365	+	0.0400
Entire_home	+	4.1326
Private_room	+	1.7334
Bronx	+	27.6436
Brooklyn	+	16.6513
Manhattan	+	26.5890
Queens	+	23.5890
Latitude_from_Manhattan	+	58.5270
Longitude_from_Manhattan	+	56.2057
Price	-	0.0017
Number_of_reviews	-	0.0521
Days_since_last_review	-	0.0017

$R^2 = 0.076$

All the features are significant (p<0.05)

→ Random Forest Regressor

R^2 training data = 0.95

R^2 test data = 0.83

Importance of features:

	importance
longitude_from_Manhattan	0.319510
availability_365	0.228667
price	0.175664
latitude_from_Manhattan	0.118571
minimum_nights	0.083392
days_since_last_review	0.032386
number_of_reviews	0.030769
Manhattan	0.004621
Entire_home	0.003581
Brooklyn	0.001453
Queens	0.001009
Private_room	0.000373
Bronx	0.000003

Conclusion

- ➔ Model predicting the number of travelers who will stay in an Airbnb in New York City using Random Forest Regressor.
- ➔ The most important features are **the longitude of the Airbnb from the center of Manhattan, the availability of the Airbnb during the year in days, the price of the Airbnb for one night and the latitude of the Airbnb from the center of Manhattan.**
- ➔ Recommendations:
 - The most important criteria to host more travelers is to **offer a price that is not too high** and to **keep his accommodation as available as possible** throughout the year.
 - For those who wish to buy a house or an apartment to convert it to Airbnb, we can recommend that they should choose **south Manhattan and northwest Brooklyn.**