Project Report

# GitHub URL

<https://github.com/GwenealB/airbnb_analysis_project>

# Abstract

This Project is made to analyze data received from Open Airbnb from any year. For now, the data is limited to 2019, 2020, and 2021 @September of each of the corresponding year.

This project features reusability of class and functions for future data provided from Open Airbnb, and allows multiple way to consume the data, that is through JSON, CSV, JSON API, or RDMS using SQLite.

# Introduction

Using Open Airbnb’s data is easy and intuitive to use, not to mention an abundance of materials on the internet that uses this Dataset with their own unique twists on their implementation gives more ideas on how to implement and work with the data given from the dataset.

# Dataset

The dataset AB\_NYC\_(Year).csv is a dataset which contains a list of Airbnb units scattered around New York city. This dataset is copied from <http://insideairbnb.com/get-the-data.html> and <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data> website. My justification of using this dataset is because of the availability of the dataset, and abundance of resources that uses, references and implements algorithm to process this dataset, easing development process for my level of skill.

# Implementation Process

Firstly, I created the git repository and published it to GitHub immediately. Having done that, I analyzed the best folder structure to store the dataset and where to place the script due to importing headache.

I decided to put scripts that would be run at the very top of the folder structure, and the rest is inserted into their respective folder (“data” and “database” for datasets and SQLite Database, “airbnb\_analysis” for project’s helper functions/class, “output” for any output file generated from the project, for example, chart images,)

I then created Python virtual environment to avoid local package conflict.

I installed required python packages (pandas, requests, matplotlib, seaborn, NumPy) then procced with the task.

The first thing I noticed that have to be done is to process the dataset and turn it into Pandas’ DataFrame, and with the project requirement requiring me to be able to process it from multiple sources, I created “airbnb\_analysis/consumer.py” to store my dataset consumer classes.

I created the classes for each type I am capable to consume (SQLite, JSON/API, CSV) so I created each class for each type.

I then created “airbnb\_analysis/analysis.py” file to process the DataFrame and make use of it to make charts, based on the project requirement.

I created AirbnbAnalyser class to store the DataFrame object, whilst bundling it with methods that can be used to make use of said DataFrame object, for example:

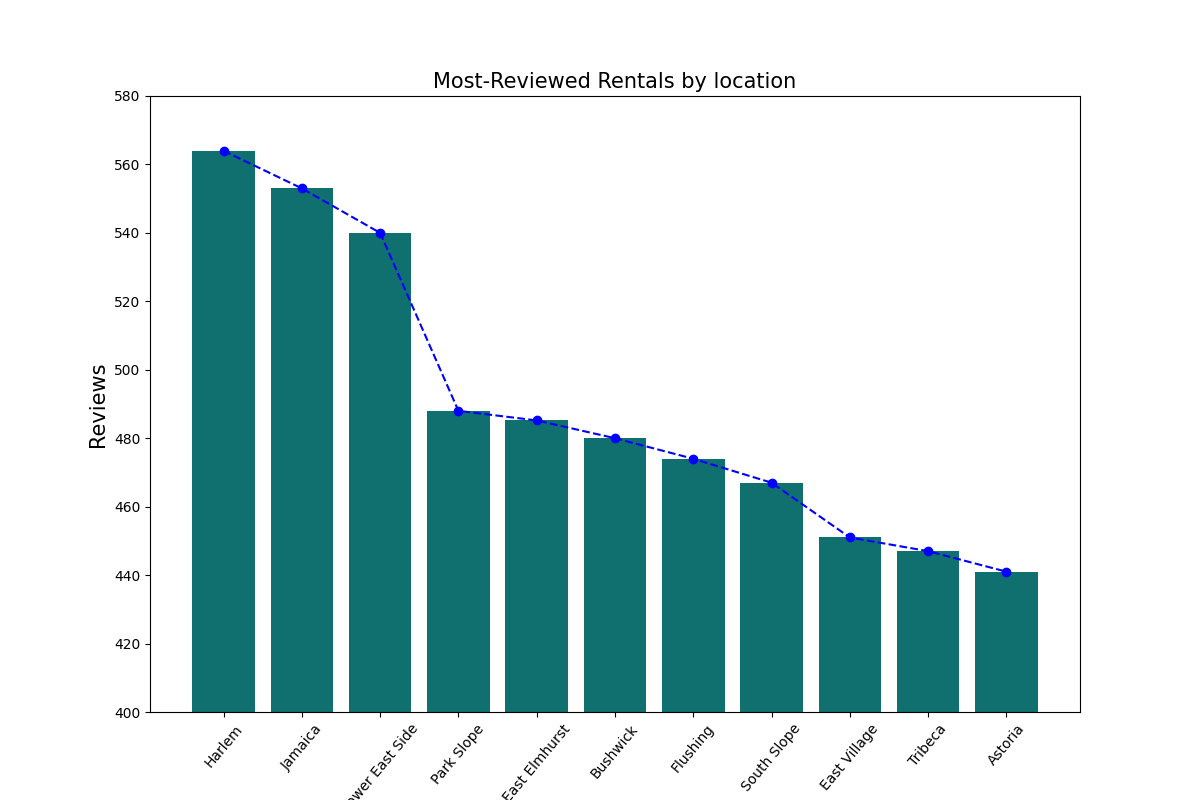
* A method to create Room Type Pie Chart.
* Airbnb unit distribution based on neighborhood.
* Unit Price Histogram.
* Price Bar Chart based on Location.
* Most Reviewed count per location.
* And lastly, a price intensity heatmap.

I then created respective python file dedicated to outputting each of those charts to PNG format, and create respective Python Jupyter file to present the data better through Jupyter notebook, for each corresponding year.

# Results

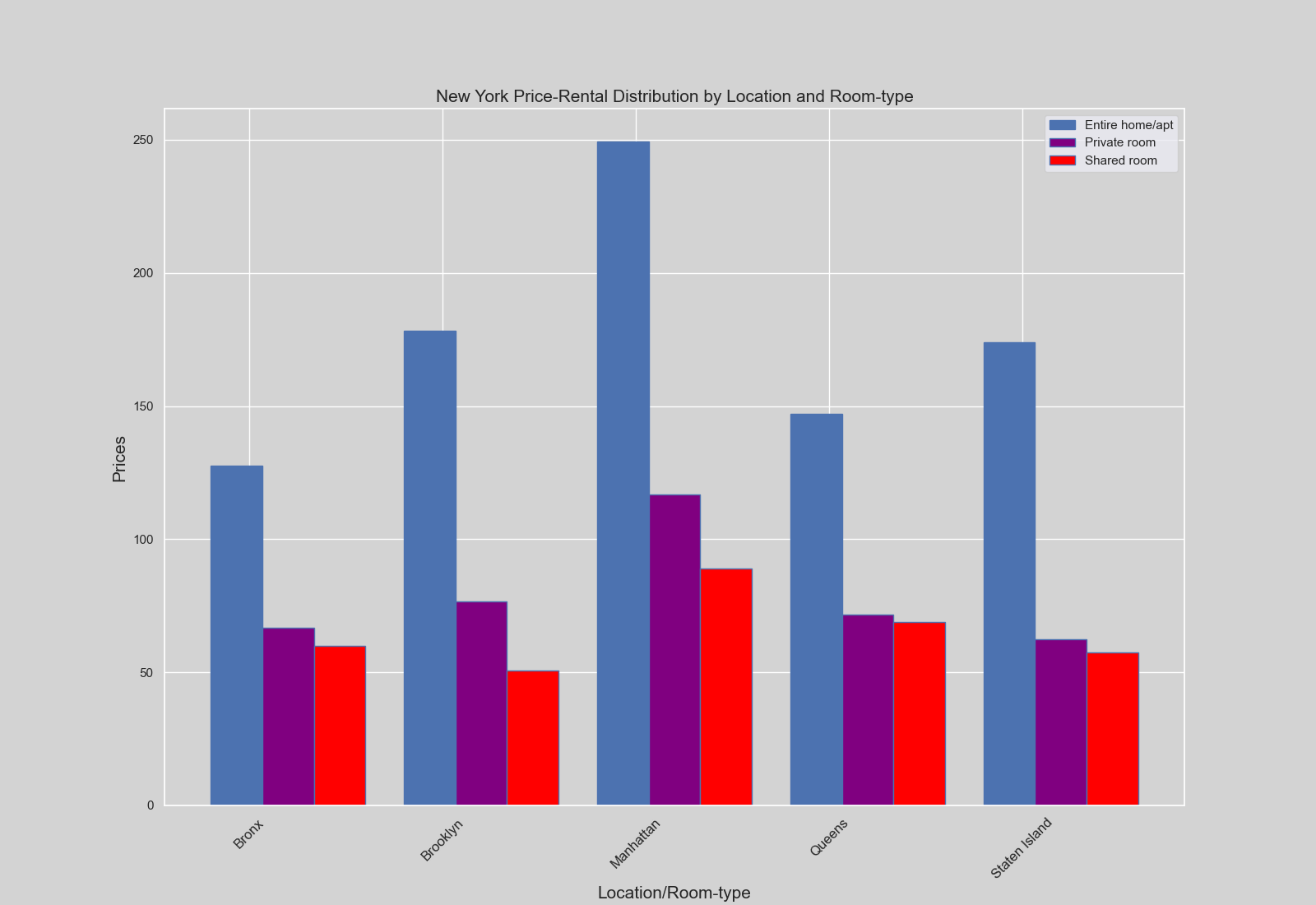
## 2019

The result described below are results for the 2019 Dataset.



Most Reviewed count per location in 2019, New York City  
Chart is peaked at ~565’s at Harlem and is at its lowest at ~440 in Astoria

Notable data: Jamaica ~550 and Sewer East Side ~535, with sudden drop of 50 review at Park Slope

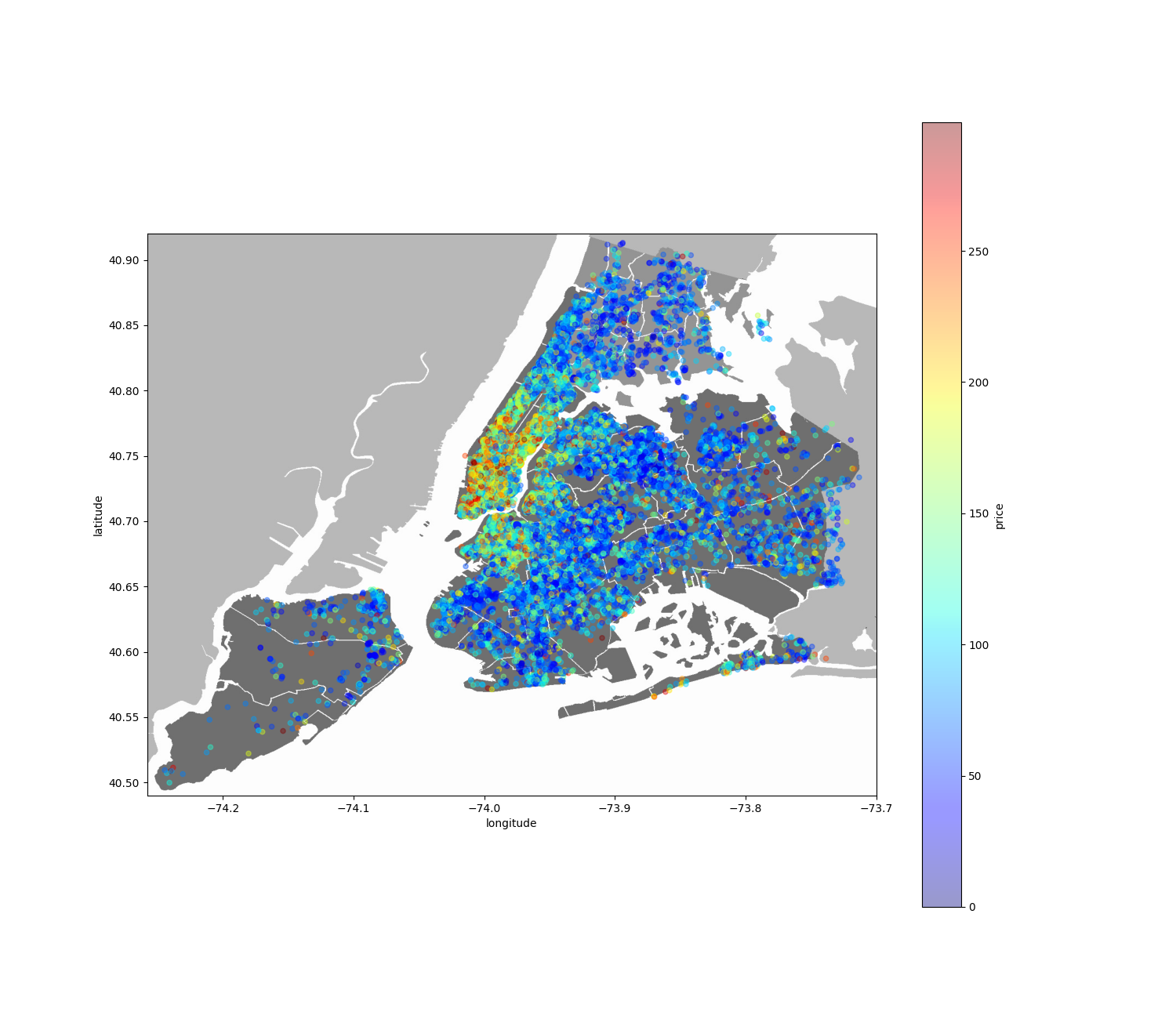


2019 New York City Airbnb Unit Price distribution by Location and room type.

On average Price peaked at 250 for Entire Home or Apartment, ~120 for Private Room, and ~85 for Shared room in Manhattan, meanwhile for lowest Entire Home/ Apartment is at Bronx for~125, Private Room is at Staten Island for ~60 and Shared room at Brooklyn for exactly 50 on average



New York City Price-Rental Distribution, highest paying rent is at ~10000 as the most paying rent, at ~100 or lower for more than 300 bookings. It can be estimated that 500 is the best price markup to set as rent rate per rent number.

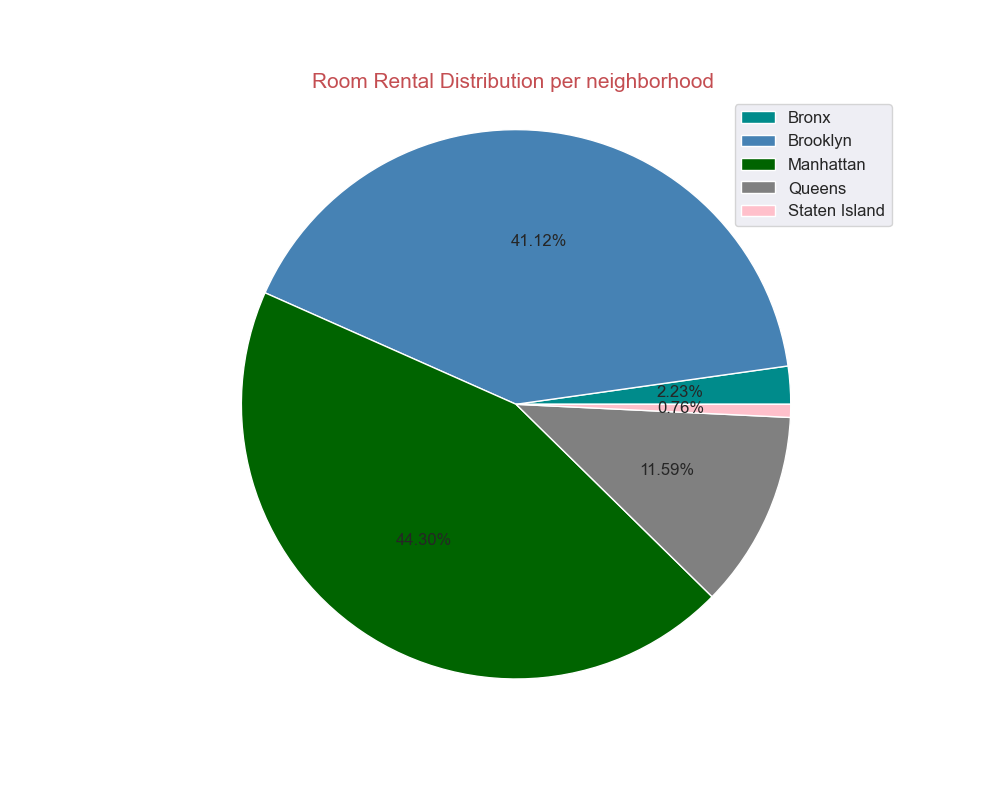


New York City Airbnb Unit Price heatmap (2019).

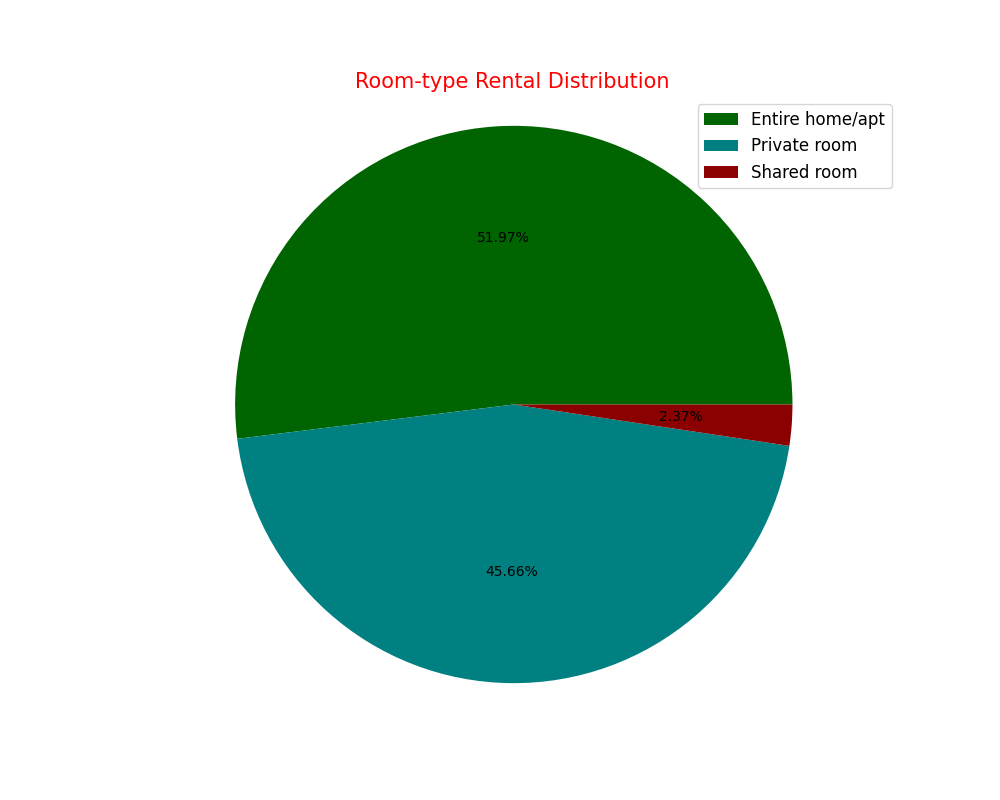
Here you can see most unit are concentrated at the center of New York City, Manhattan, with most high priced there, meanwhile the concentration on the outskirts is way scarce than in Manhattan.

Units in Staten Island can be seen as scarcer than on Queens.

Unit Price in Brooklyn can be seen a slightly higher than other part of the city besides Manhattan, which may be caused due to price effects from Manhattan.



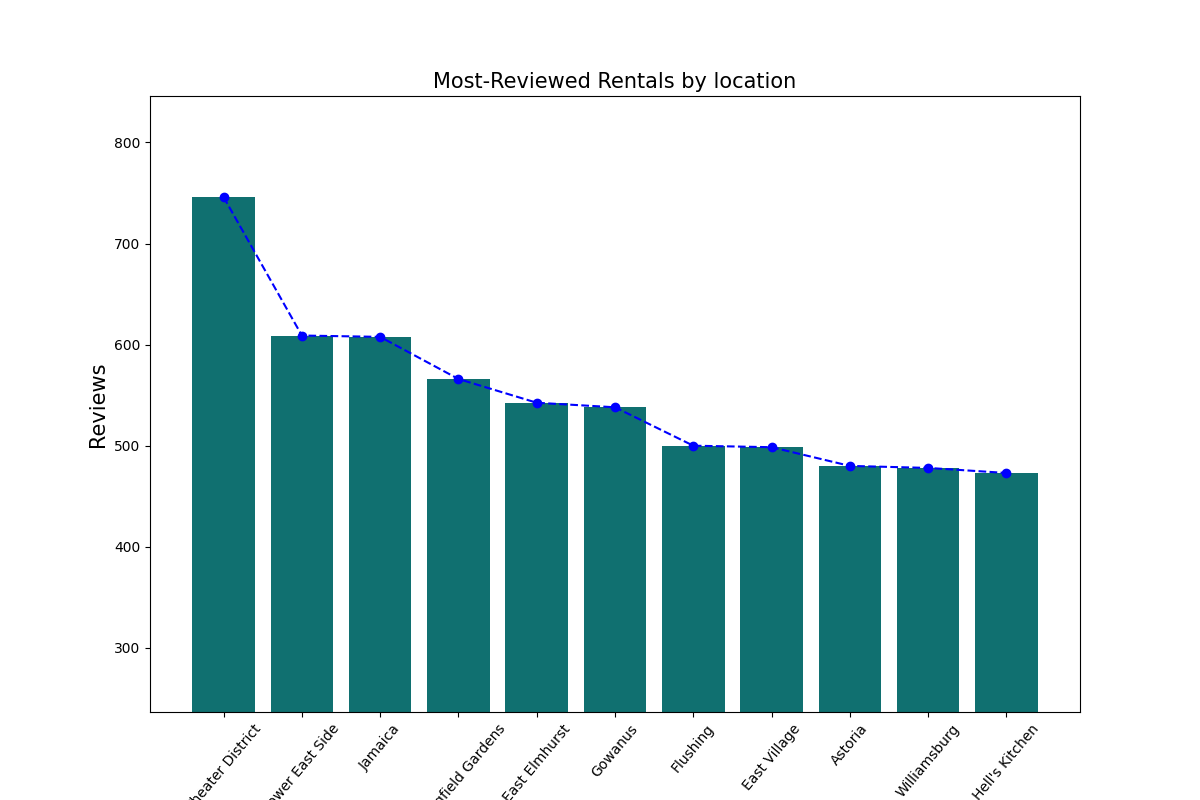
Airbnb Unit share per location per neighborhood in 2019, highest at Manhattan at 44.40% share, lowest at Staten Island at 0.76% share



Room types distribution of Airbnb units in New York City in 2019, highest is Entire Home/Apartment for more than half, and lowest is Shared room at measly 2.37%

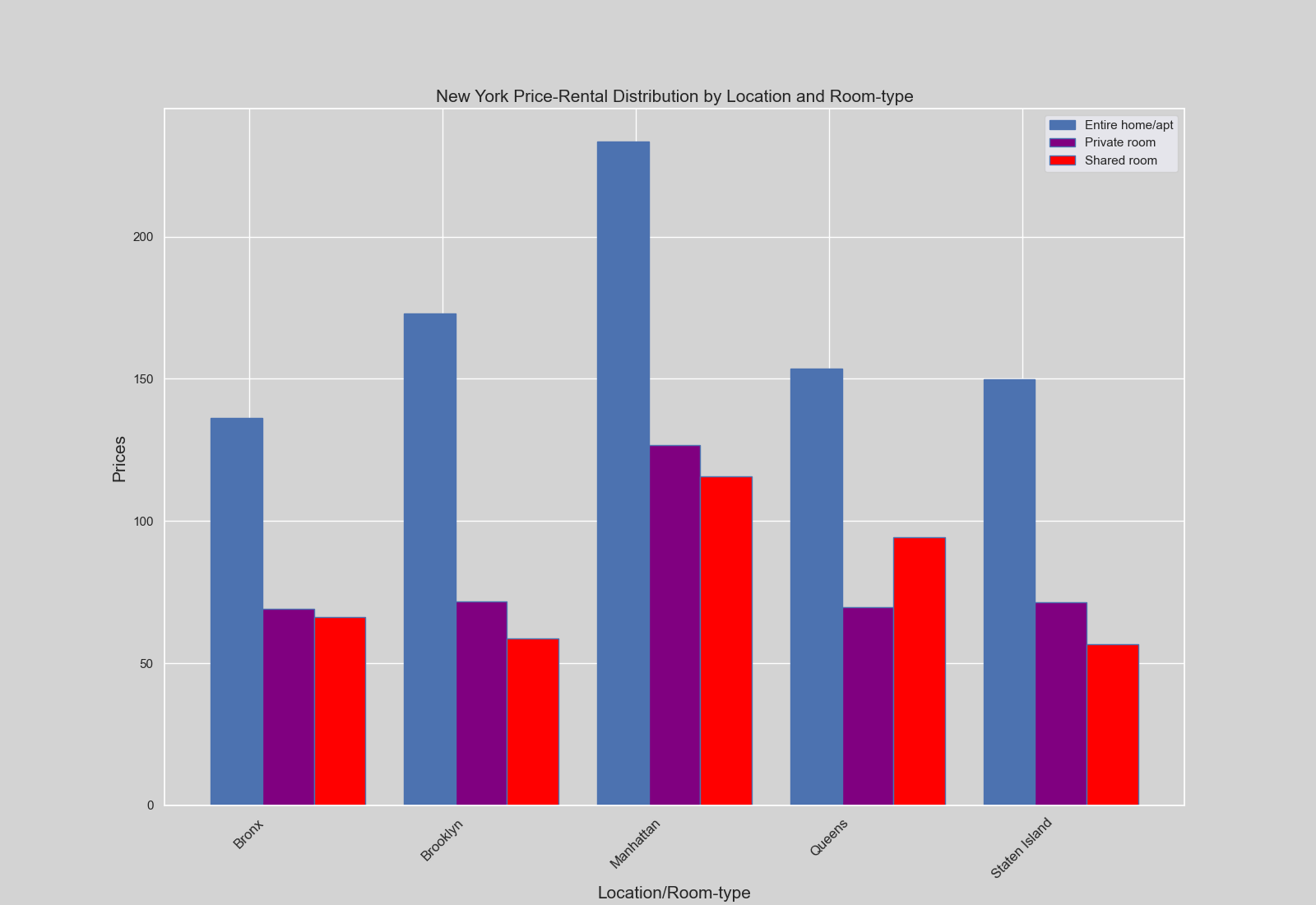
## 2020

The result described below is for the year 2020



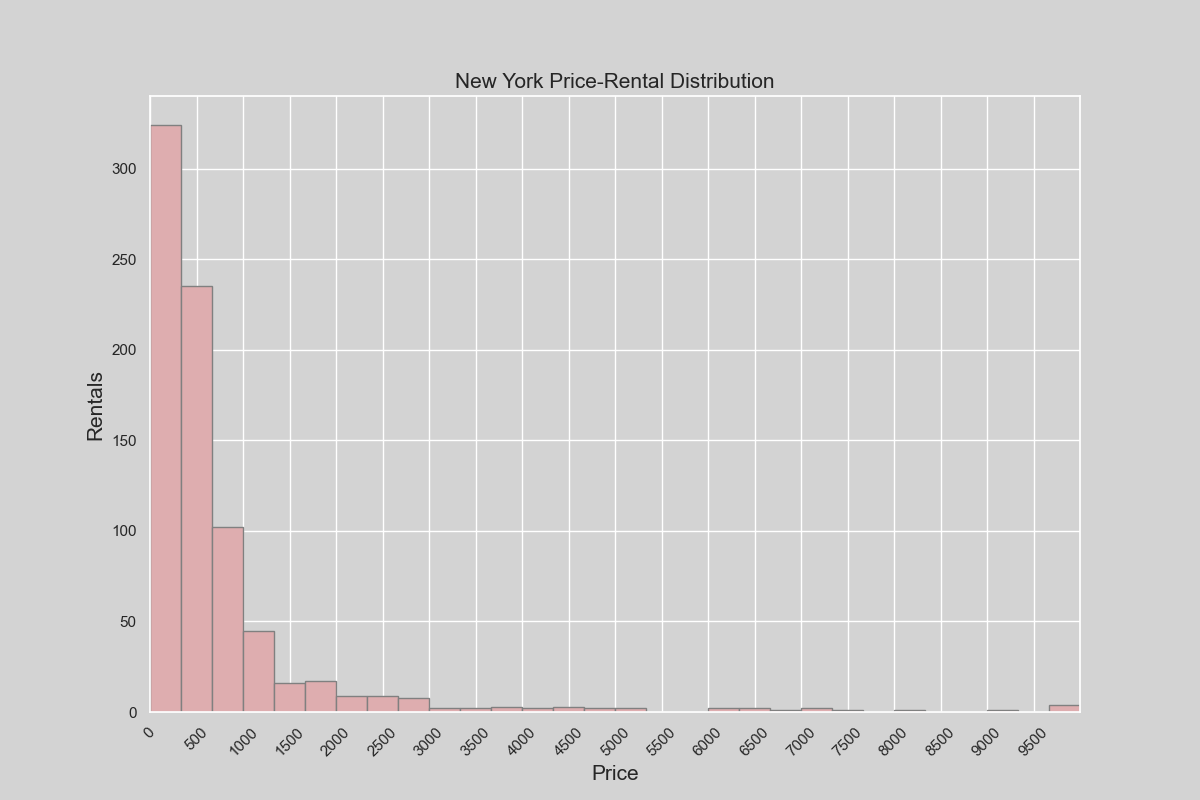
Most Reviewed count per location in 2020, New York City  
Chart is peaked at ~750’s at Theater District and is at its lowest at ~490 in Hell’s Kitchen

Notable data: A drop of ~150 reviews making Sewer East side and Jamaica at 600 reviews

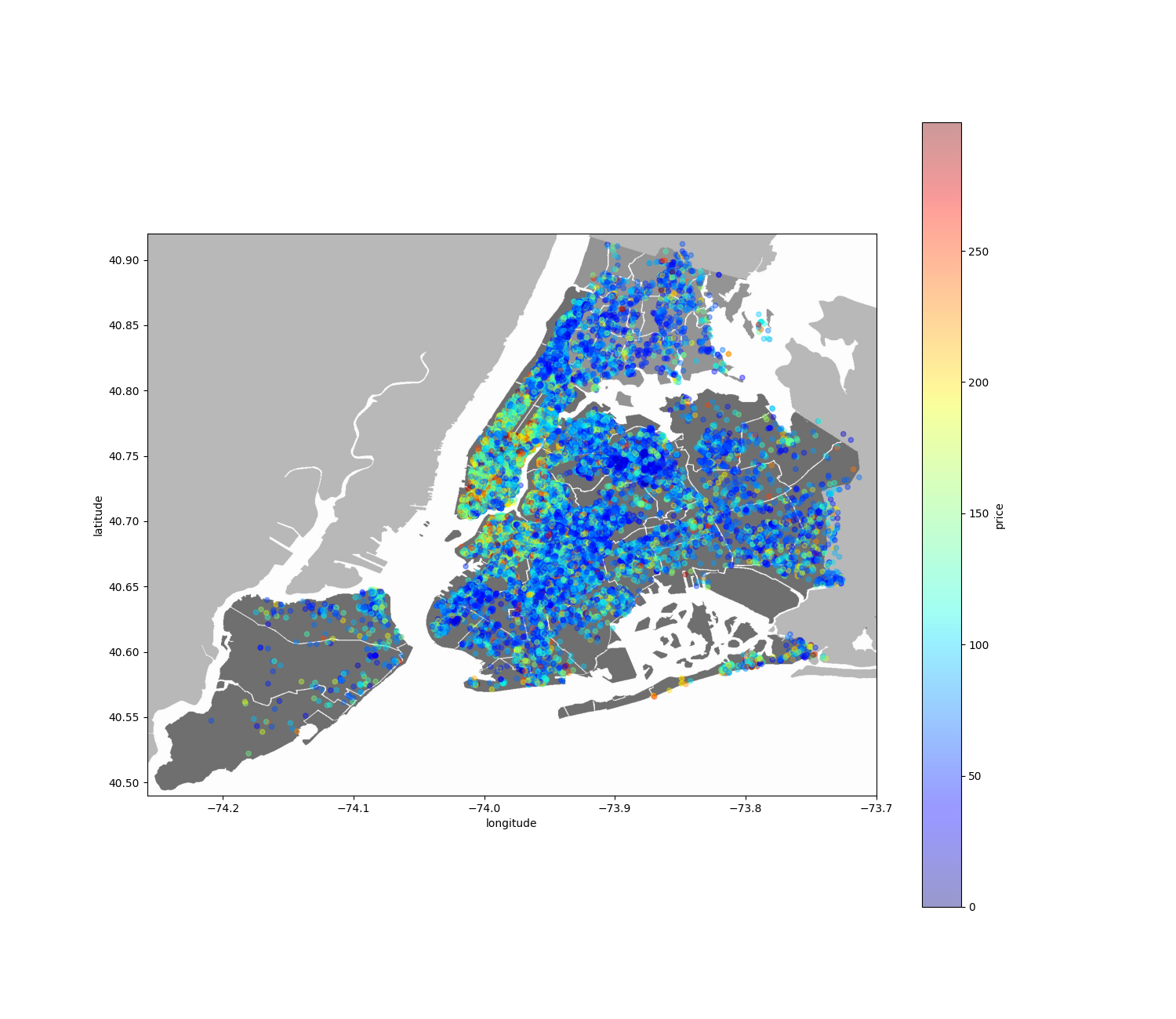


2020 New York City Airbnb Unit Price distribution by Location and room type.

On average Price peaked at 240 for Entire Home or Apartment, ~125 for Private Room, and ~115 for Shared room in Manhattan, meanwhile for lowest Entire Home/ Apartment is at Bronx for~130, Private Room is same for Bronx, Brooklyn, Queens and Staten Island for ~65 and Shared room at Brooklyn for ~55 on average



2020 New York City Price-Rental Distribution, highest paying rent is still at ~10000 as the most paying rent with slightly more rent, at ~100 or lower for more than 340 bookings. It can still be estimated that 500 is the best price markup to set as rent rate per rent number.



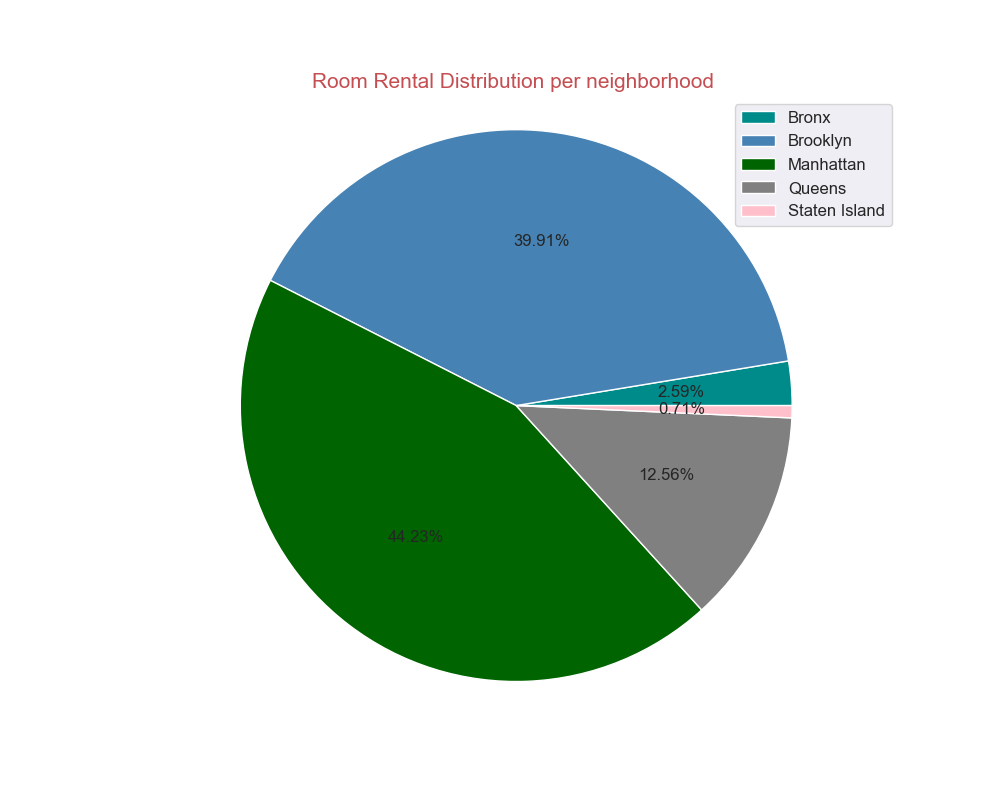
New York City Airbnb Unit Price heatmap (2020).

Here you can see most unit are concentrated at the center of New York City, Manhattan, with most high priced there, meanwhile the concentration on the outskirts is way scarce than in Manhattan.

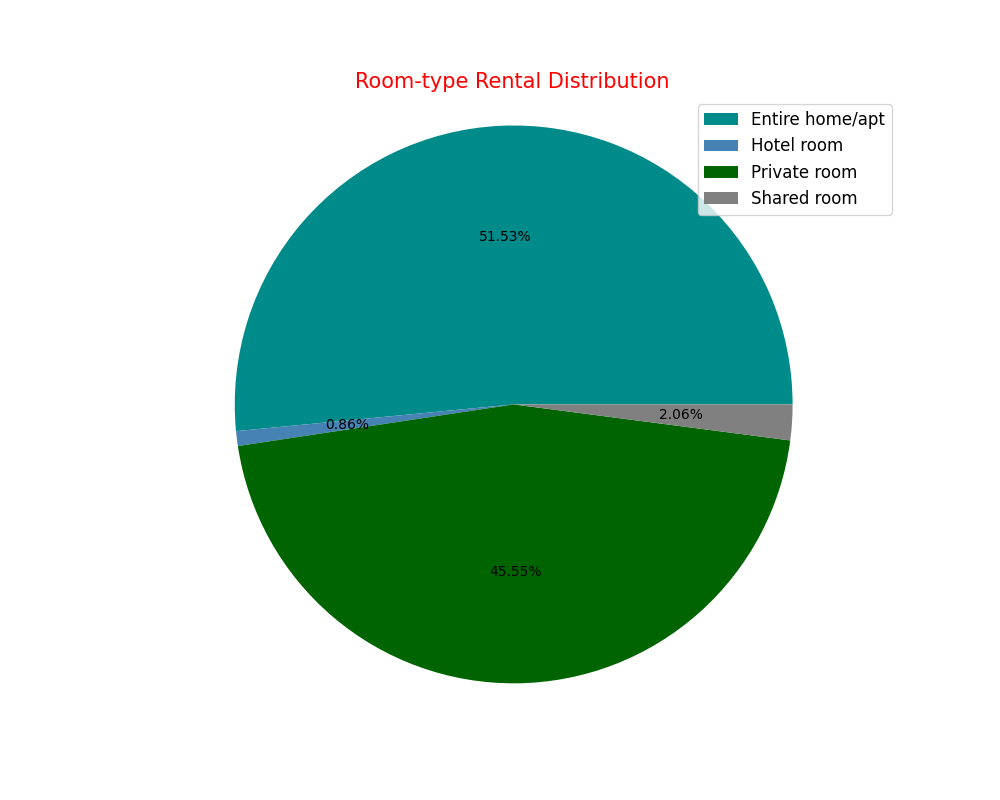
Compared to last year, there seems to be a drastic price fall, which can be pointed to the pandemic.

Units in Staten Island can still be seen as scarcer than on Queens. No signs of noticeable development.

Unit Price in Brooklyn can be seen a slightly higher than other part of the city besides Manhattan, which may be caused due to price effects from Manhattan. But compared to last year there seems to be price fall, again, which can be attributed to the COVID pandemic.

\

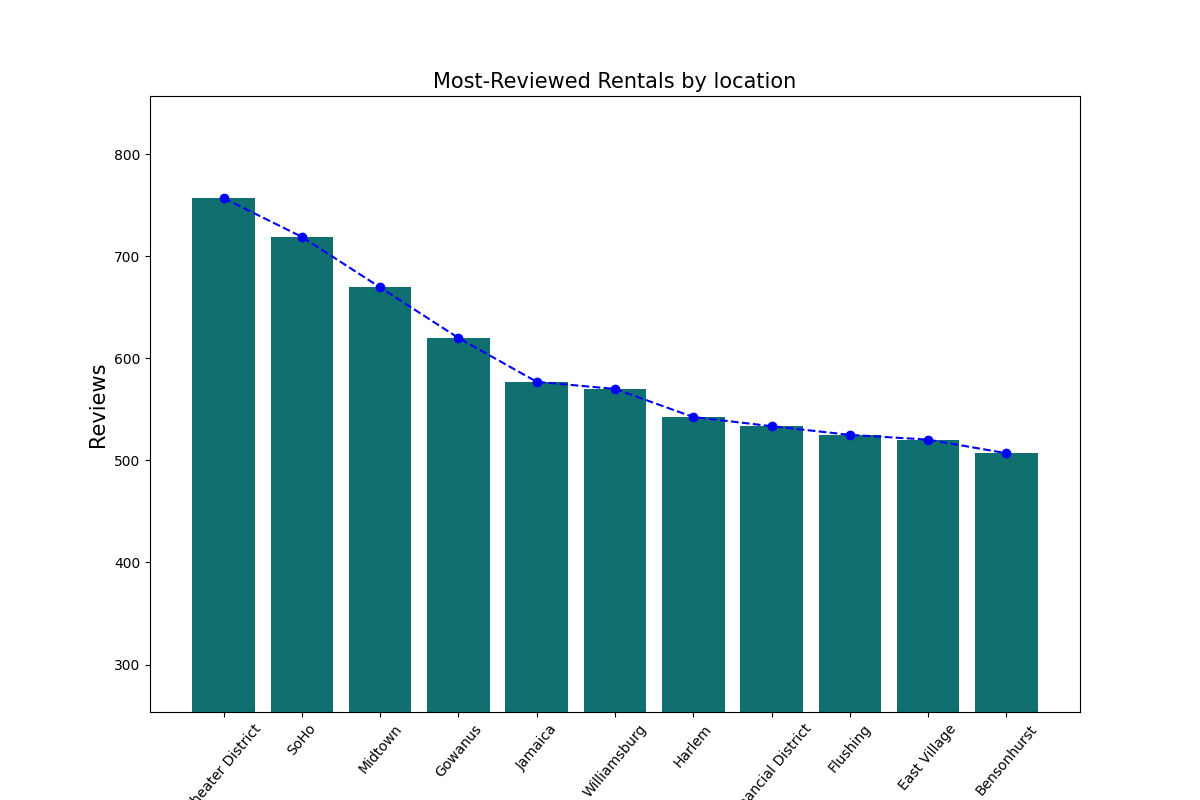
Airbnb Unit share per location per neighborhood in 2020, highest at Manhattan at 44.23% (-0.17%) share, lowest at Staten Island at 0.71% share, lower than 0.05% than last year



Room types distribution of Airbnb units in New York City in 2020, highest is still Entire Home/Apartment for more than half, and lowest is Hotel room at measly 0.86%

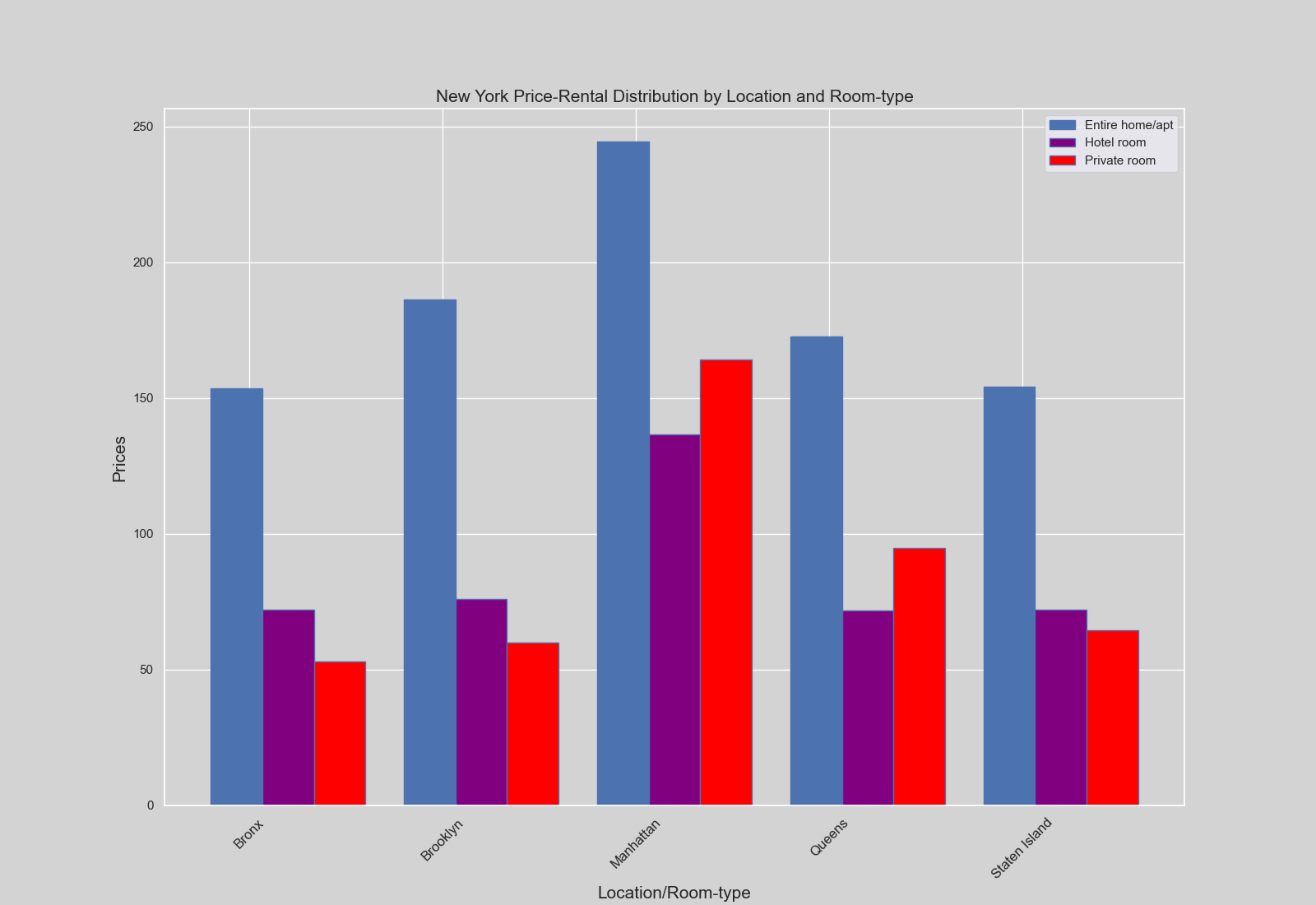
## 2021

The result below is for the year 2021



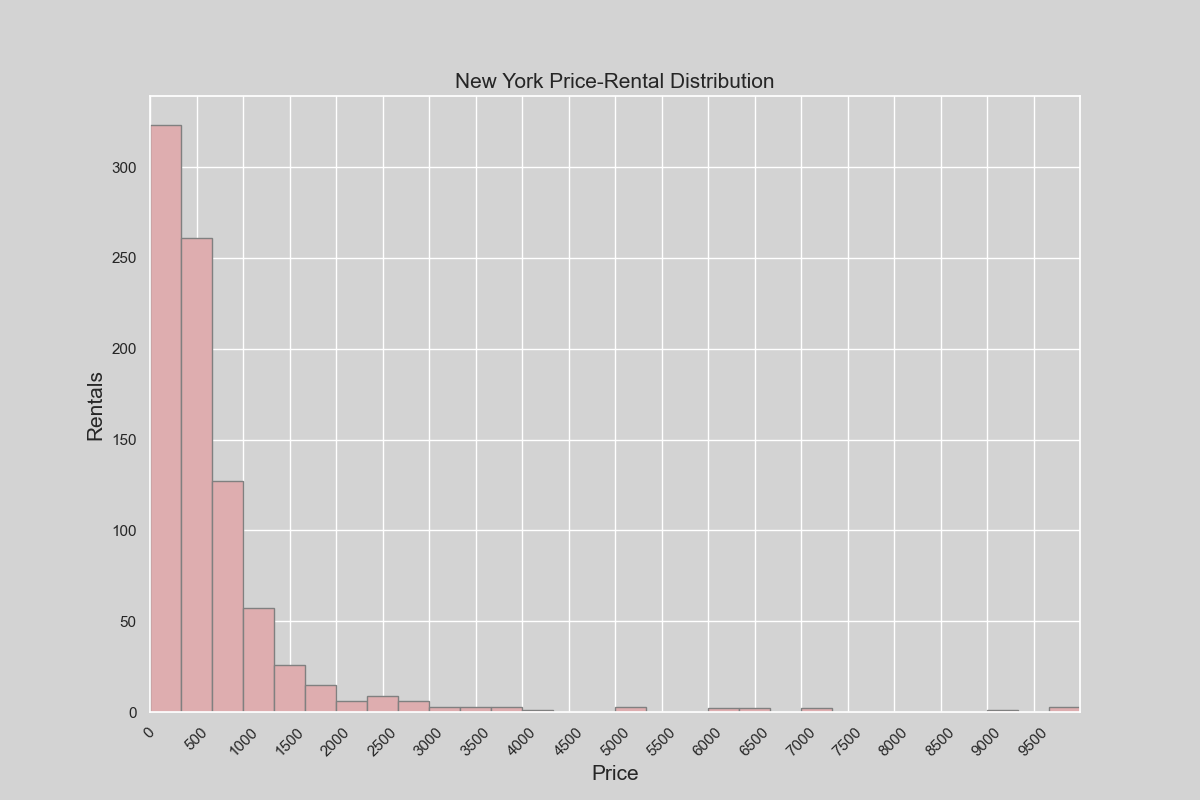
Most Reviewed count per location in 2021, New York City  
Chart is peaked at ~760’s at Theater District and is at its lowest at ~520 in Bensohurst

Notable data: Jamaica at 580 reviews

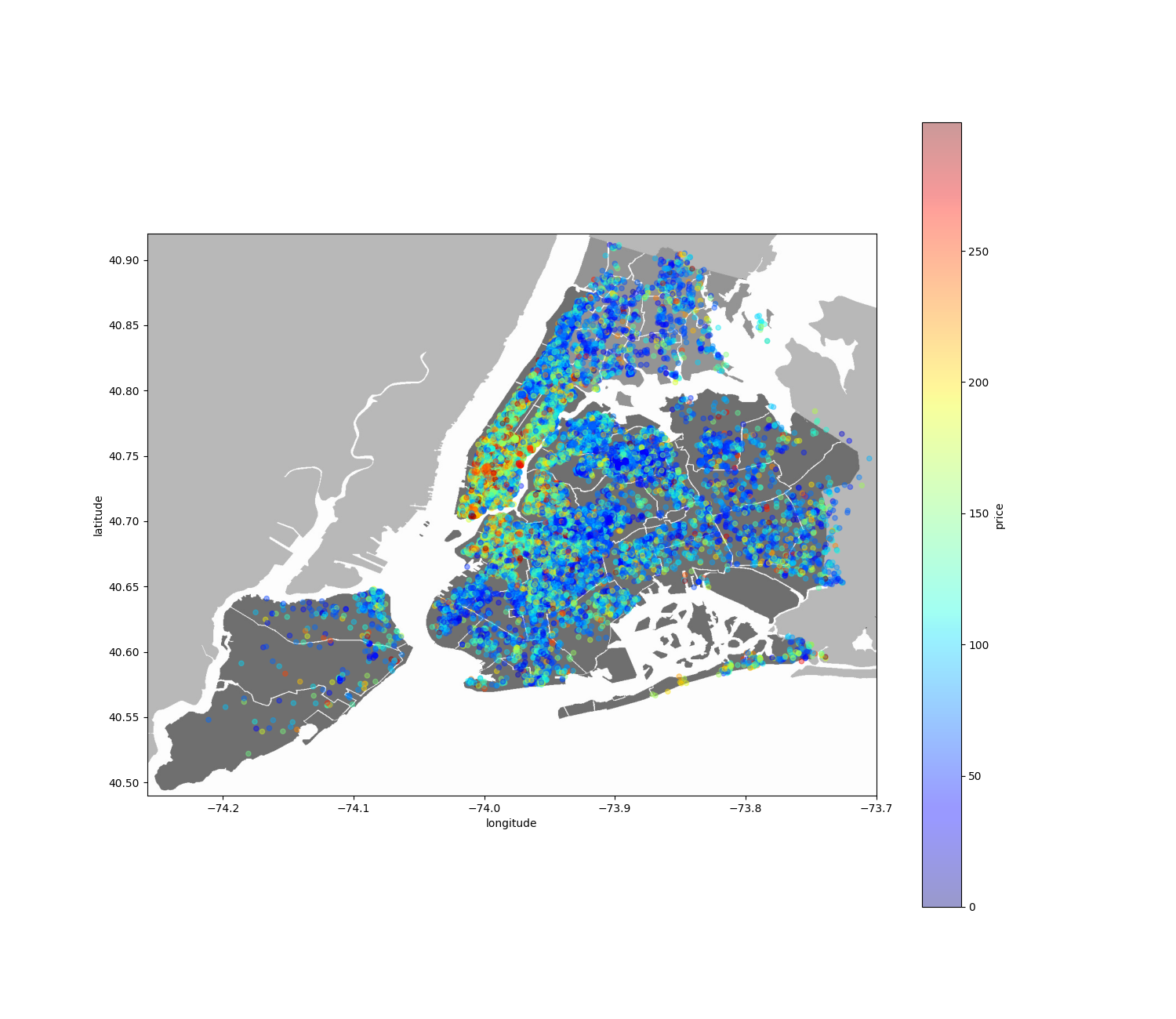


2021 New York City Airbnb Unit Price distribution by Location and room type.

On average Price peaked at almost 250 for Entire Home or Apartment, ~160 for Private Room, and ~145 for hotel room in Manhattan, meanwhile for lowest Entire Home/ Apartment is at Bronx and Staten Island for~155, Private Room is at Bronx, ~70 and Shared room at Bronx for ~55 on average



2021 New York City Price-Rental Distribution, highest paying rent is still at ~10000 as the most paying rent with less rent than last year, at ~100 or lower for more than 340 bookings. It can still be estimated that 500 is the best price markup to set as rent rate per rent number.



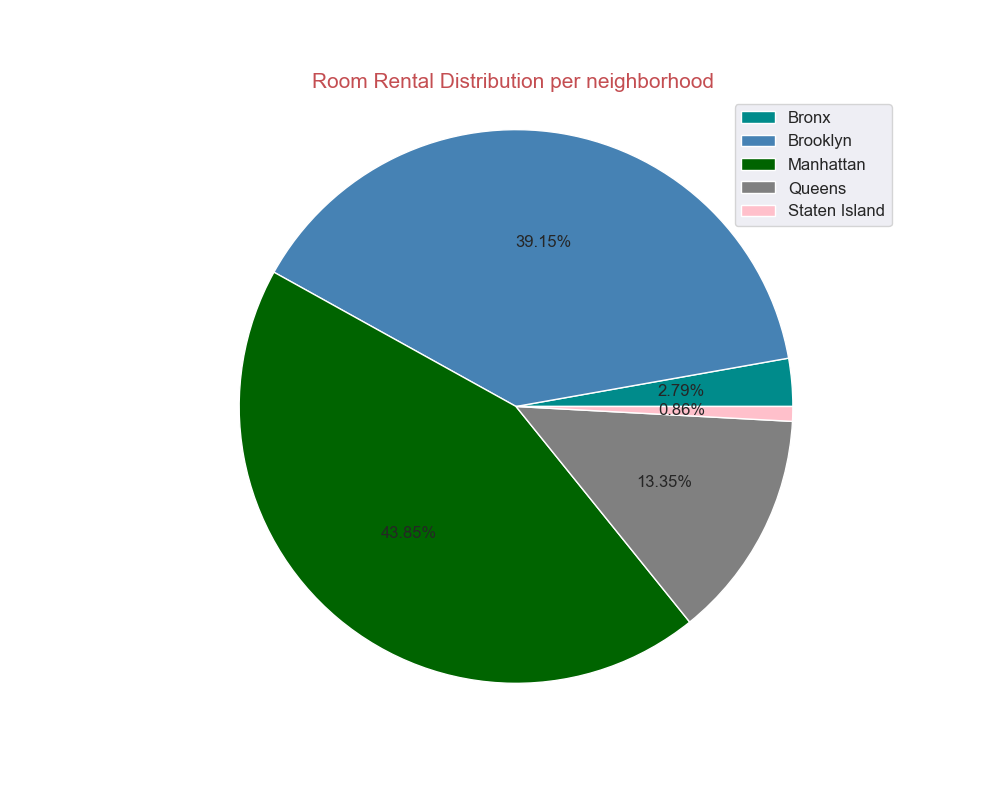
New York City Airbnb Unit Price heatmap (2021).

Here you can see most unit are concentrated at the center of New York City, Manhattan, with most high priced there, meanwhile the concentration on the outskirts is way scarce than in Manhattan.

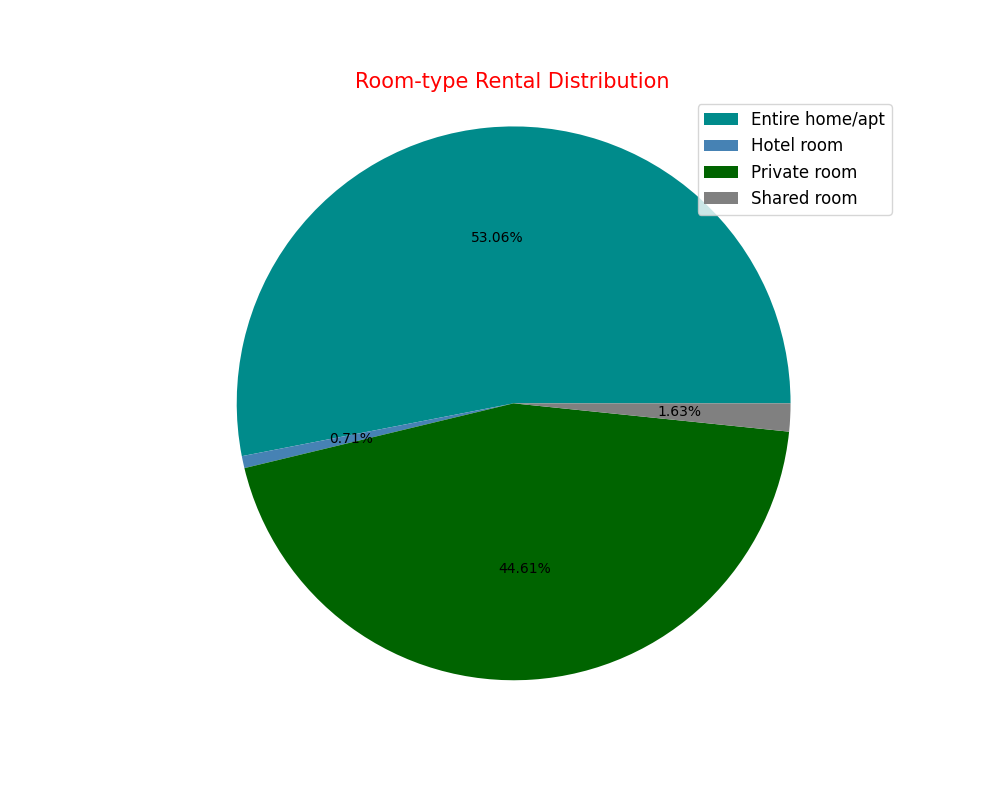
Compared to last year, there seems to be a price rise, which can be pointed to the pandemic calming.

Units in Staten Island can still be seen as scarcer than on Queens. No signs of noticeable development.

Unit Price in Brooklyn can be seen a slightly higher than other part of the city besides Manhattan, which may be caused due to price effects from Manhattan. But compared to last year there seems to be price rise, again, which can be attributed to the COVID pandemic losing momentum on economy.



Airbnb Unit share per location per neighborhood in 2021, highest at Manhattan at 43.85% (-0.37%) share, lowest at Staten Island at 0.86% share, higher 0.15% more than last year



Room types distribution of Airbnb units in New York City in 2020, highest is still Entire Home/Apartment for more than half (53%), and lowest is Hotel room at measly 0.71%

# Insights

* Insight 1: Most Reviewed Airbnb units is most likely at Theater District and Jamaica for the fact that they both constantly have a high number of reviews over the years.
* Insight 2: Most expensive units are resided in Manhattan, while the cheapest is in Bronx.
* Insight 3: Best Price/Rent you can set your unit for is around $500 per unit.
* Insight 4: Setting up a unit in Manhattan might prove to be highly competitive, and it’s suggested to set up more on the outskirts if you set in without any strategy.
* Insight 5: Brooklyn and Manhattan are packed with Airbnb unit, if combined make more than 75% of total units in New York City.
* Insight 6: The market is overflowed with private accommodation, mostly an entire home or apartment, followed next with private room. Shared room pales in comparison.
* Insight 7: There is a new Room type introduction in 2020, Hotel Room.
* Insight 8. Chart indicates price fall on 2019-2020, which can be attributed to the COVID-19 pandemic.
* Insight 9: The effect of the COVID pandemic has slightly subsided as seen on the heatmap chart of 2020-2021
* Machine Learning: Using more accurate data, I can create Airbnb unit price prediction based on the number rents and more accurate and monthly data instead of yearly. I can also create Aibnb unit spread prediction. I would use regression based moel for my machine learning due to the data being continuous and the predicion needed is continuous in nature too.

# References

* Kaggle: Dgomonov - New York City Airbnb Open Data; Airbnb listings and metrics in NYC, NY, USA (2019) <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>
* Get the Data - Inside Airbnb. Adding data to the debate. @New York City, New York, United States <http://insideairbnb.com/get-the-data.html>
* Towards Data Science: Data Analysis of New York Rentals using Python <https://towardsdatascience.com/airbnb-rental-analysis-of-new-york-using-python-a6e1b2ecd7dc>