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.

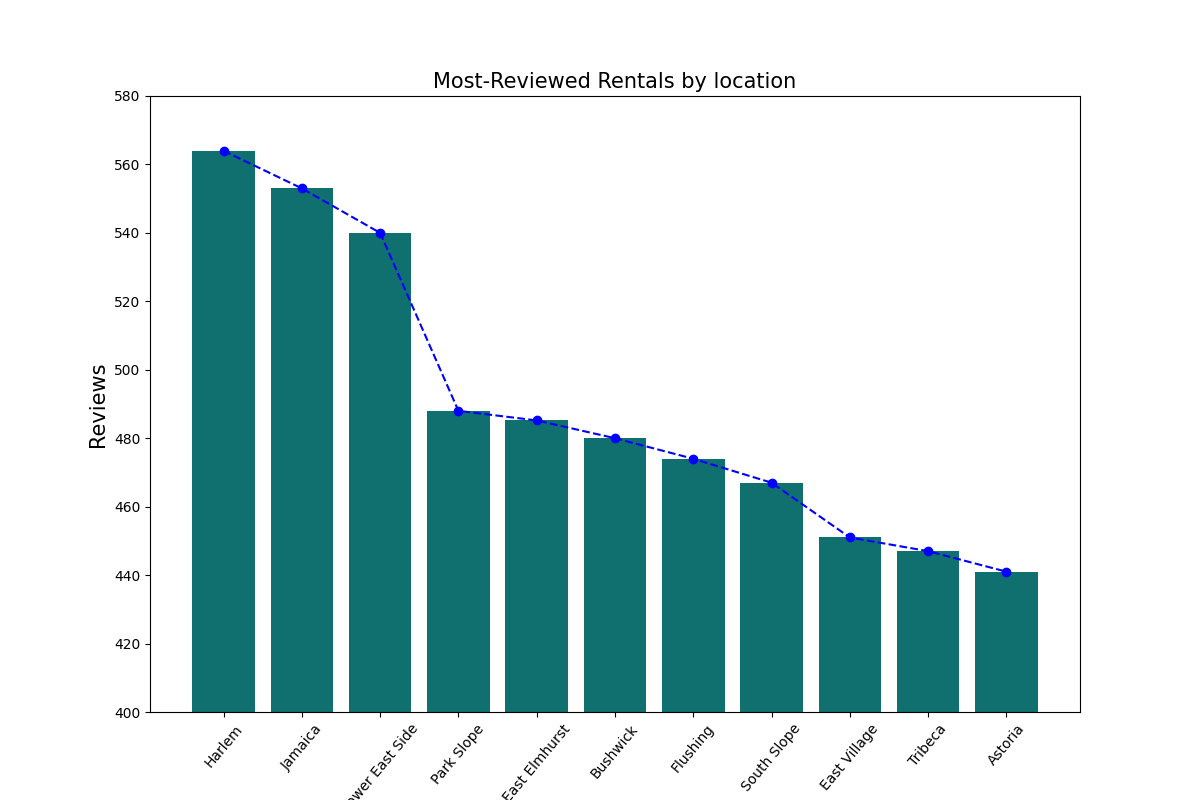
I created AirbnbAnalyser class to store the DataFrame object, whilst bundling it with methods that can be used to make use of said DataFrame object, for example:

* 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.

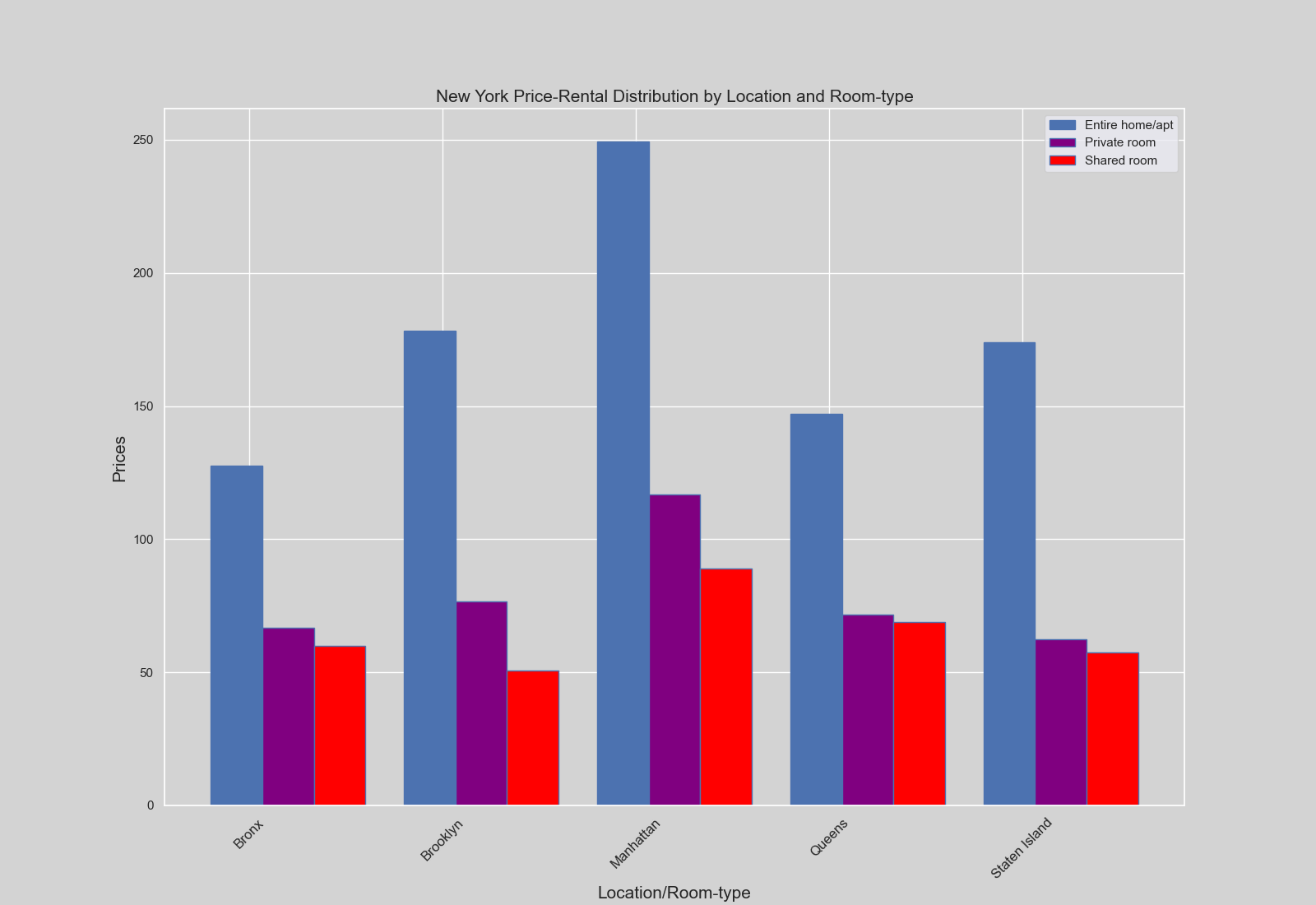
I then created respective python file dedicated to outputting each of those charts to PNG format, and create respective Python Jupyter file to present the data better through Jupyter notebook, for each corresponding years.

# Results

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 it’s lowest at ~40 in Astoria

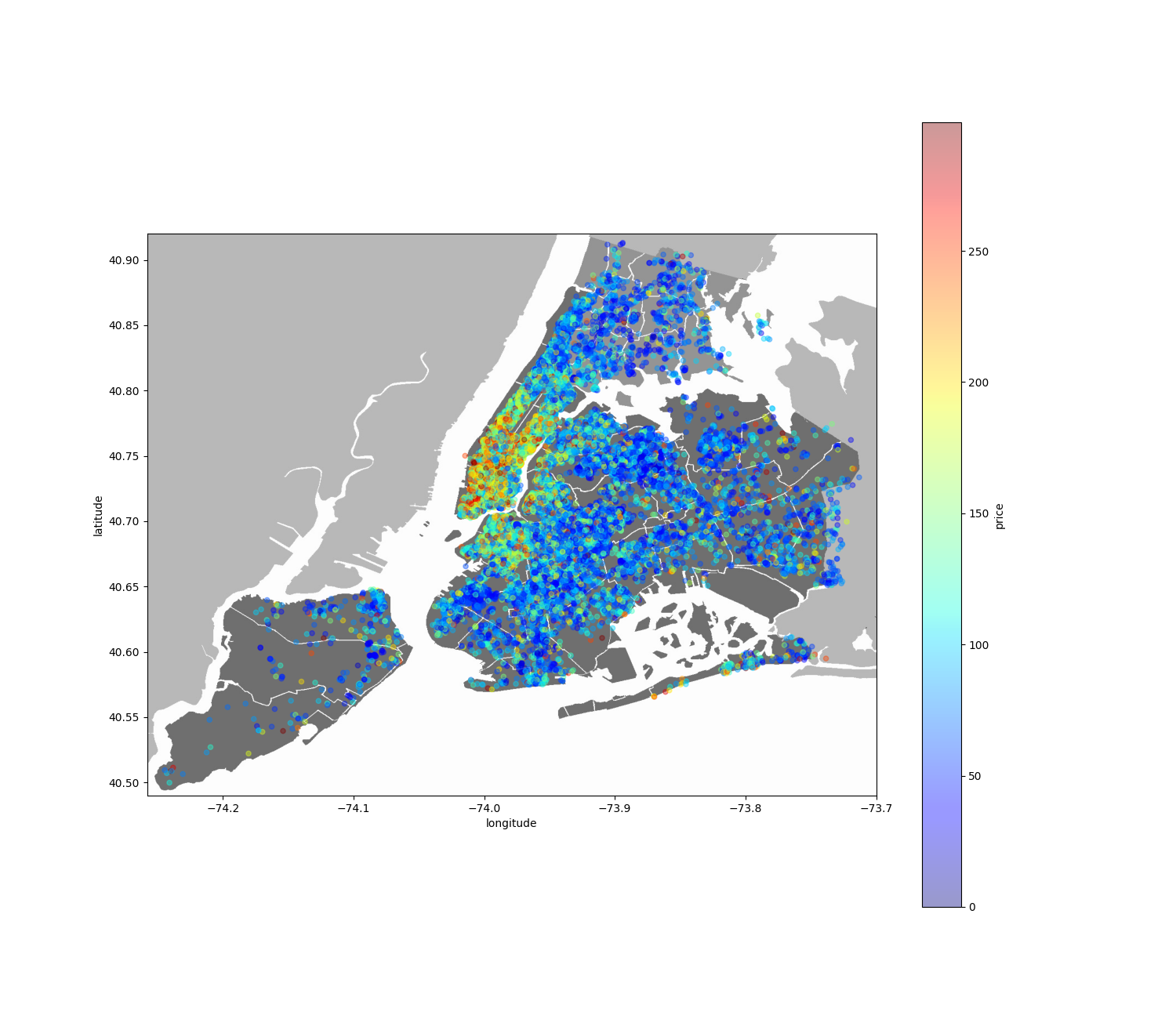


New York City Airbnb Unit Price distribution by Location and room type.

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

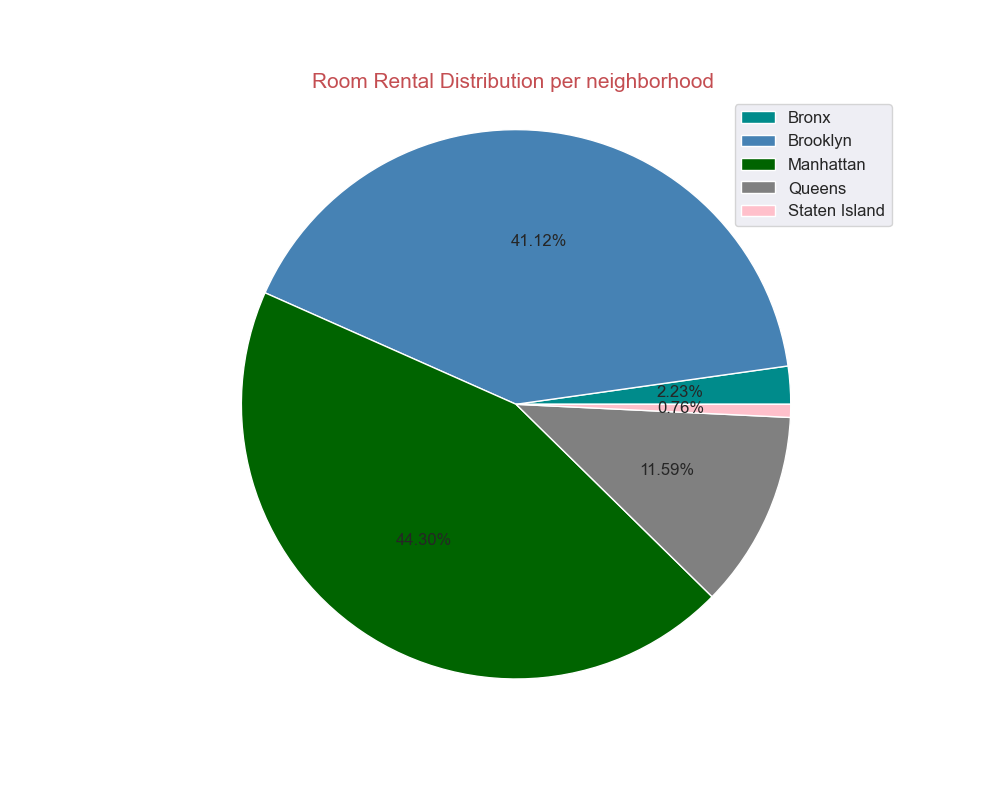


New York City Price-Rental Distribution, highest paying rent is at ~10000 and most paying rent, obviously at ~100 or lower for more than 300 bookings.

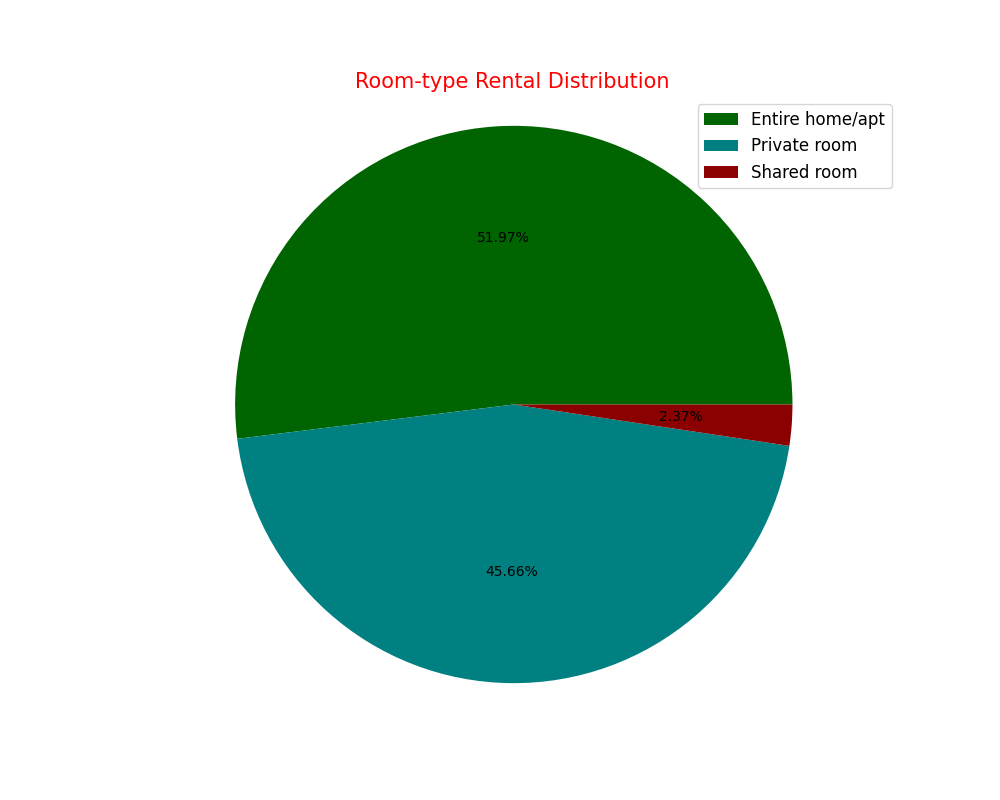


New York City Airbnb Unit Price heatmap.

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



Airbnb Unit share per location per neighborhood, 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, highest is Entire Home/Apartment for more than half, and lowest is Shared room at measly 2.37%

# Insights

* Insight 1:

# 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>