Final Project Report

Jacob Gulan

**1 INTRODUCTION**

Airbnb is a popular mobile and web application that offers users the ability to easily rent out their place to other people at the user’s own specified price. For my final project, I used a public dataset from Kaggle [1] that provides open-source data on Airbnb listings in New York City. This dataset provides various features like latitude, longitude, room type, price, availability, reviews per month, and other various features. The features focused on for this project are latitude, longitude, price, and room type. The goal of this project is to study and find a relationship between location, price, and room type.

**2 APPROACHES**

Before applying the data to construct the visualization, I pre-processed the data. Using Python, I pre-processed the data into a CSV file that would then be imported into my Unity project along with a C# script. In the preprocessing step, I removed most of the features besides latitude, longitude, price, and room type. These features are the ones I wanted to visualize the relationships of. Along with removing excess features, I also normalized the data and randomly sampled about 250 entries out of the 49,000+ entries to produce a better visualization. Two outliers were also removed from the data to provide a clearer visualization. Using this data and the Unity Editor, I devised two visualizations to model the data that will assist in studying relationships within the data.

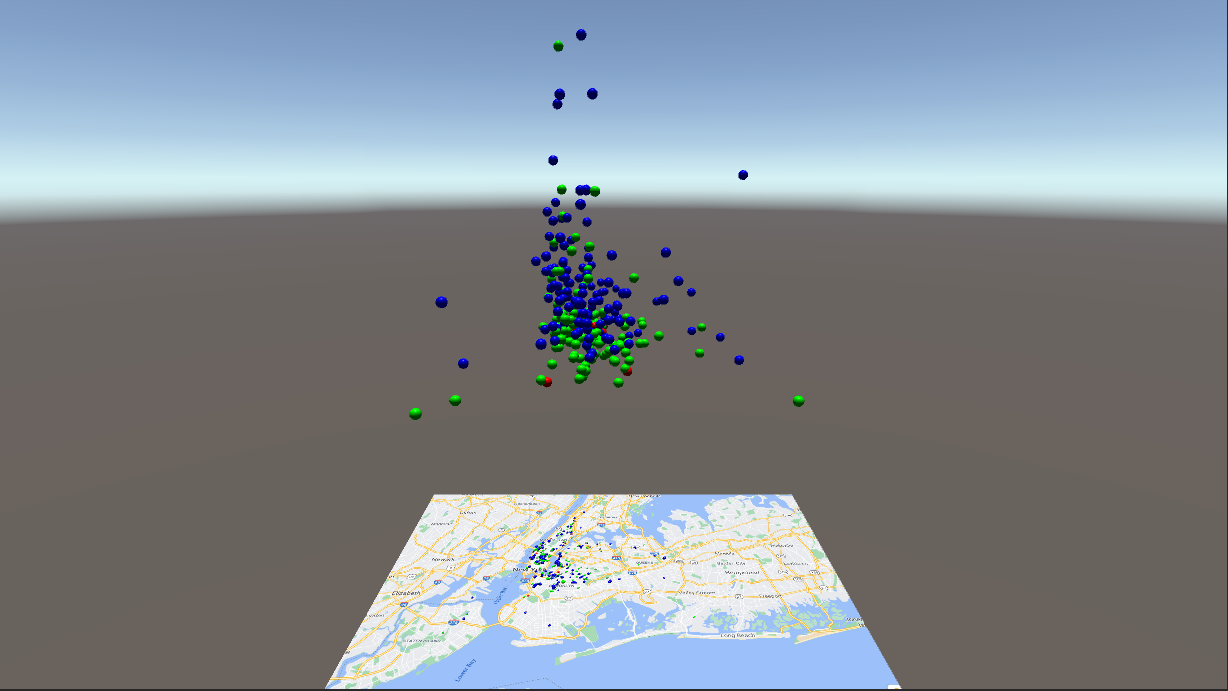
The first visualization I created was a scatterplot. To construct this plot, I implemented a C# script that reads a CSV file containing the normalized data. From this data, a sphere is created for every entry according to its price, room type, latitude, and longitude. The spatiality of the spheres is determined by the price, latitude, and longitude. The longitude denoted by the x-value and the latitude denoted by the z-value is used to create spatial differences between the data to visualize the data easier. Spheres that appear close to each other in terms of horizontal distance indicate that the entries are close together in real life as well. The price, denoted by the y-value, is used to show the expenses of each entry. The higher the sphere appears on the scatterplot the more expensive the listing is and the lower the sphere is the less expensive the listing is. To establish clarity between which sphere falls under what room type the spheres are colored according to their room type. Green is a private room also denoted as 1 in the dataset, blue is the entire home/apartment also denoted as 2 in the dataset, and red is a shared room also denoted as 3 in the dataset. Depending on the color of the sphere, it was either named SphereGreen, SphereBlue, or SphereRed to ease the process of implementing interaction features.

The second visualization I created was a geo-visualization that also models the relationship between latitude, longitude, price, and room type. In this visualization, the data is plotted onto a map of New York City respective to its actual real-world location by using the latitude and longitude provided by the dataset. Similar to the previous visualization, the spheres are colored according to their room type where green is a private room, blue is the entire home/apartment, and red is a shared room. The price is used to determine the size of the sphere in which it is determined by a scale. The scale is as such: $0-100, $100-200, $200-300, $300-400, $400-500, > $500. The smallest spheres on the graph are the $0-100 listings and the largest spheres are the ones greater than $500. Again depending on the color of the sphere, it was either named SphereGreen, SphereBlue, or SphereRed to ease the process of implementing interaction features.

Now, to discuss the interaction features applied. The visualization starts off in a multi-coordinated system that displays both of the visualizations. To explore each visualization further in depth the camera may be toggled by pressing the spacebar in which the camera will change to being primarily focused on one of the visualizations. These cameras may be cycled through for an infinite amount of repetitions by pressing the spacebar. If the camera is focused on either the geo-visualization or the scatterplot visualization the spheres visualized may be reduced to only spheres of a specific category. By pressing “Z,” only the private room category will be visible. By pressing “X,” only the entire home/apartment category will be visible. By pressing “C,” only the shared room category will be visible. By pressing “R,” the visualization will return to its original state. The script achieves this by searching for the name of the objects and enabling or disabling them depending on the button pressed. Also, if scatterplot visualization is currently being viewed then by pressing “A” or “D” the camera will rotate around the scatterplot offering a better look at the spatialityof the graph. This is achieved by placing an invisible GameObject in the center of the scatterplot and having the camera rotate around that.

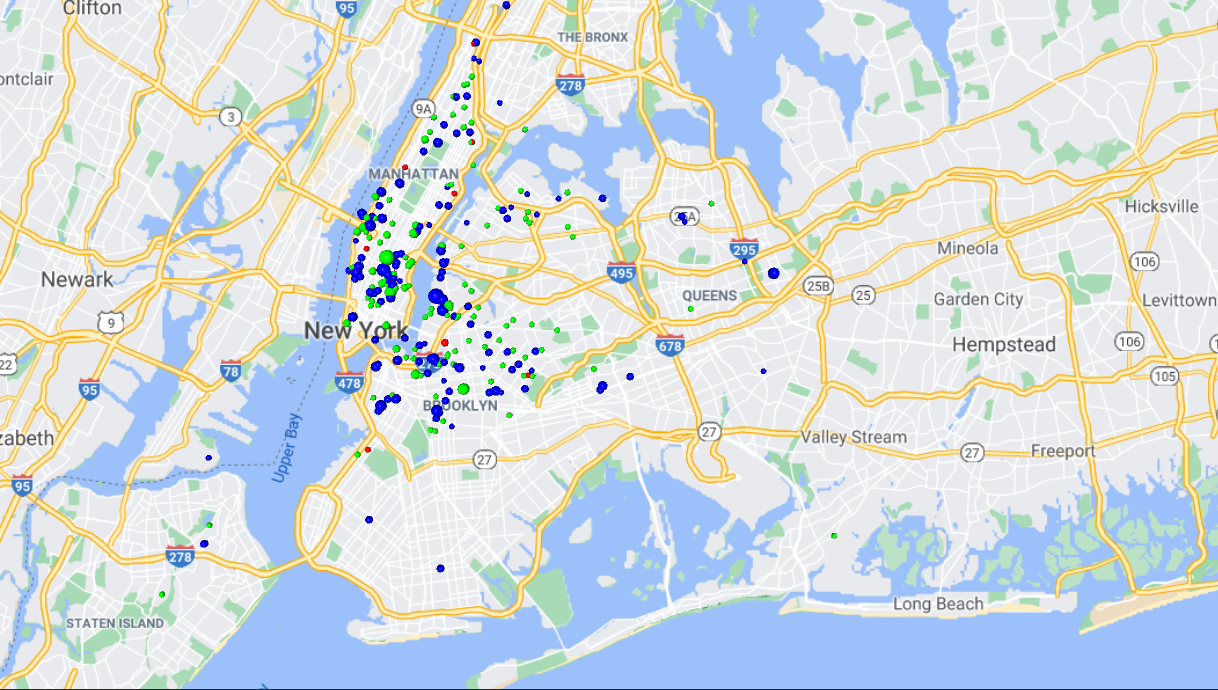
**3 RESULTS**

With the approach implemented, let’s observe the results of the visualizations. In *Figure 1,* the main camera shows the multi-coordinated system is visible upon launching the project. This provides a good first-look glance at the project, but to understand and study the relationships more let’s switch the camera.



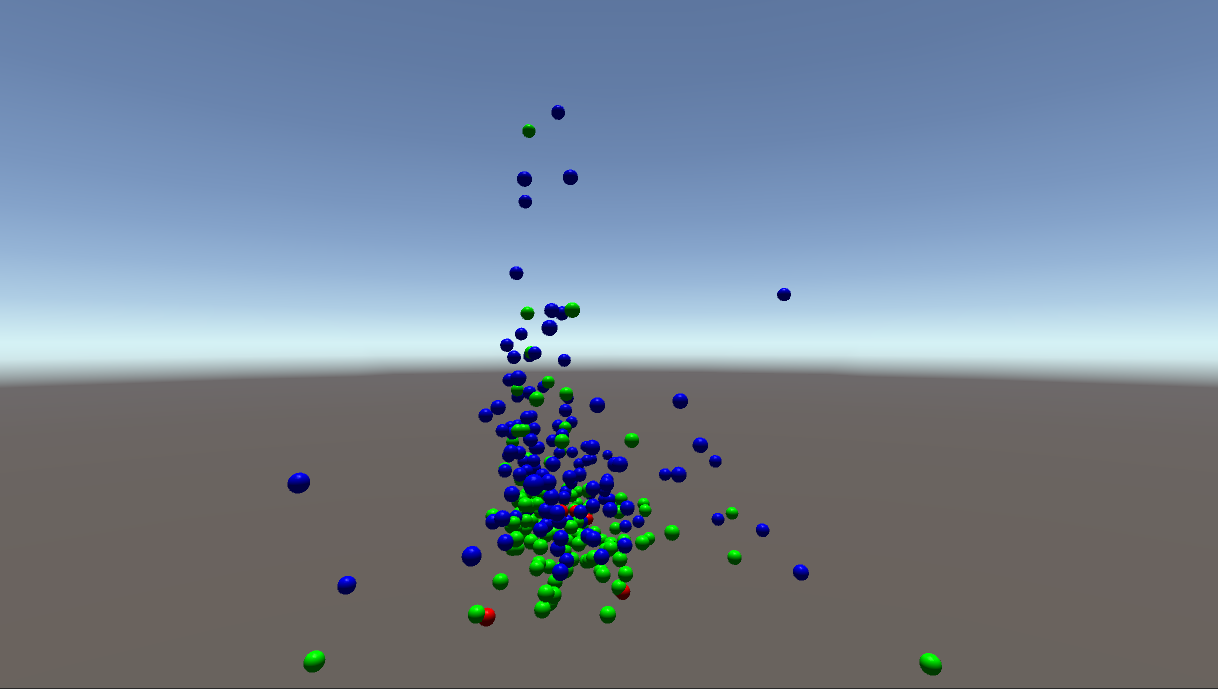
*Figure 1 - Multi-coordinated system visualization of New York City Airbnb Open Data*

The geo-visualization of the data can be seen in *Figure 2*. From the data, it reveals some interesting relationships and information. It appears that the most expensive listings from the data are centralized in the Southern Manhattan region as well as Northwest Brooklyn. Most of the listings also appear to come from the central parts of New York with few listings coming from boroughs like The Bronx, Staten Island, and Queens.

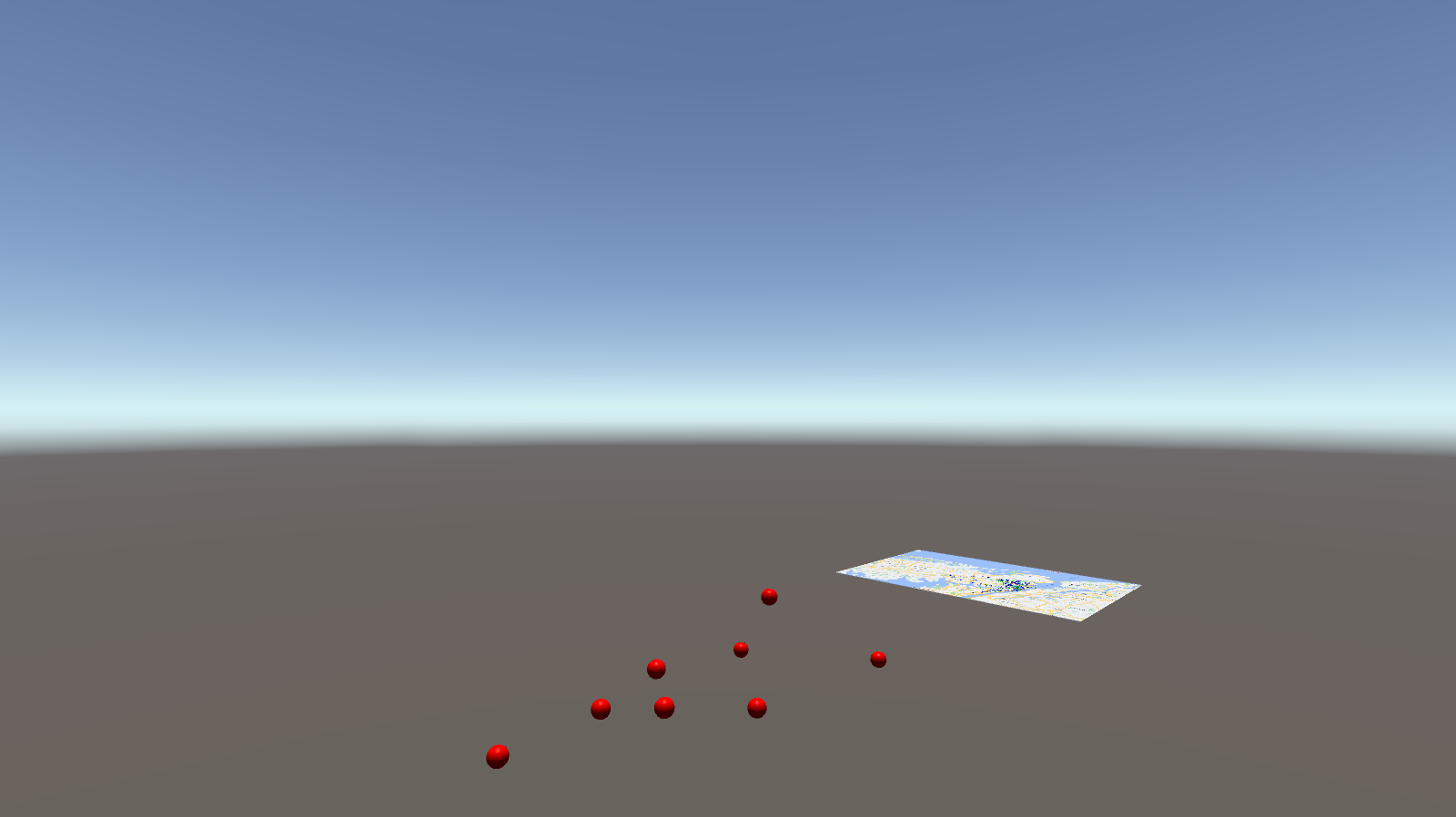


*Figure 2 - Geo-visualization of New York City Airbnb Open Data*

The scatterplot visualization is able to offer more insight into the prices of the Airbnb listings relative to the others. In *Figure 3*, the scatterplot visualization is shown. From this figure, it appears most of the listings are clustered together with relatively similar spatiality and price. It appears that the entire home/apartment category has the highest averaging price based on their sphere’s height relative to the other categories. In *Figure 4*, the camera is rotated to better visualize the spatiality of the spheres and just the category of shared rooms is visible. This figure shows that the shared room listings do not carry an expensive price tag and most of the listings are separated with an adequate amount of spatial distance indicating not many of these listings are close to each other.



*Figure 3 - Scatterplot visualization of New York City Airbnb Open Data*



*Figure 4 - Rotated camera showing scatterplot visualization of exclusively shared room listings*

**4 CONCLUSION**

The goal of this project was to study and find a relationship between location, price, and room type. From the geo-visualization, the data shows that most of the listings on Airbnb for New York City occur within the Southern Manhattan and North West Brooklyn area. These listings also appear to have a heavier price tag than listings outside of this area. From this, it would appear there’s an association between the price of the listing and its location on the map. It’s reasonable to assume that the prices for listings found in Southern Manhattan and North West Brooklyn occur due to the demand. From the scatterplot, the data shows that listings with a room type of entire/home apartment tend to have the highest average prices, private rooms follow with the second-highest average prices, and shared rooms have the lowest average price. It would be reasonable to assume that users value privacy and so listings with a room type of either a private room or entire/home apartment cost more than a shared room. These visualizations seem to strongly suggest that there are relationships between location, price, and room type in regard to the New York City Airbnb Open Data.

**5 REFERENCES**

[1] Dgomonov. “New York City Airbnb Open Data.” *Kaggle*, Kaggle, 12 Aug. 2019, <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>. Accessed 23 November 2020.