Best Airbnb Rentals for NYU Students

By Charles Hu

1. Introduction

1.1 Background

Studying at a university in the United States, while beneficial, comes with a hefty price tag. Nowhere is this more true than at New York University (NYU), where students must routinely pay over $70,000 a year to attend; $18,000 of which is for room and board (costs of living in university dormitories) [1]. It is therefore desirable for students of NYU to reduce these costs as much as possible. One way this can be achieved is by renting Airbnb's (a company that offers cheap rentals) instead of living in the university's dormitories.

1.2 Problem

In this project, we will investigate how to help reduce NYU students' financial burden by locating the best Airbnb's. This means that while it should be affordable, it also shouldn't detract from a student's university experience. Therefore, one of the factors that should be minimized is distance to the university to reduce excessive travel time to and from classes. The Airbnb should also ideally be in close proximity to a plethora of venues. And finally, we must choose Airbnb's that have a reasonable price.

1.3 Interest

The primary stakeholders of this project, NYU students, may attempt to look for cheap Airbnb accommodations close to the university by themselves. However, due to the large amount of Airbnb rentals in New York City - close to 50,000 - students may be faced with a choice overload, and hence choose an option that is less than ideal. By narrowing down the options for NYU students to a smaller subset of more preferable Airbnb's, we can help them make a smarter choice tailored to their preferences.

2. Data

2.1 Data Sources

The dataset that we will base our research on will be the ‘New York City Airbnb Open Data’ Dataset from Kaggle [2]. This dataset contains data for 48,895 Airbnb listings in New York City in 2019. There are 16 features describing each listing, shown below in Table 1.

Table 1

NYC Airbnb Dataset Feature Description

Index	Feature Name	Description	datatype
0	id	Unique listing id for each listing	int64
1	name	Name of the listing	object
2	host_id	Unique host id	int64
3	host_name	Name of the host	object
4	neighbourhood_group	Burough	object
5	neighbourhood	Area within the borough	object
6	latitude	Latitude coordinates	float64
7	longitude	Longitude coordinates	float64
8	room_type	Type of the listing space	object
9	price	Price in dollars per night	int64
10	minimum_nights	Amount of nights minimum per stay	int64
11	number_of_reviews	Number of reviews	int64
12	last_review	Date of last review	object
13	reviews_per_month	Number of reviews per month	float64
14	calculated_host_listings_count	Amount of listings per host	int64
15	availability_365	Number of days when listing is available for booking	int64

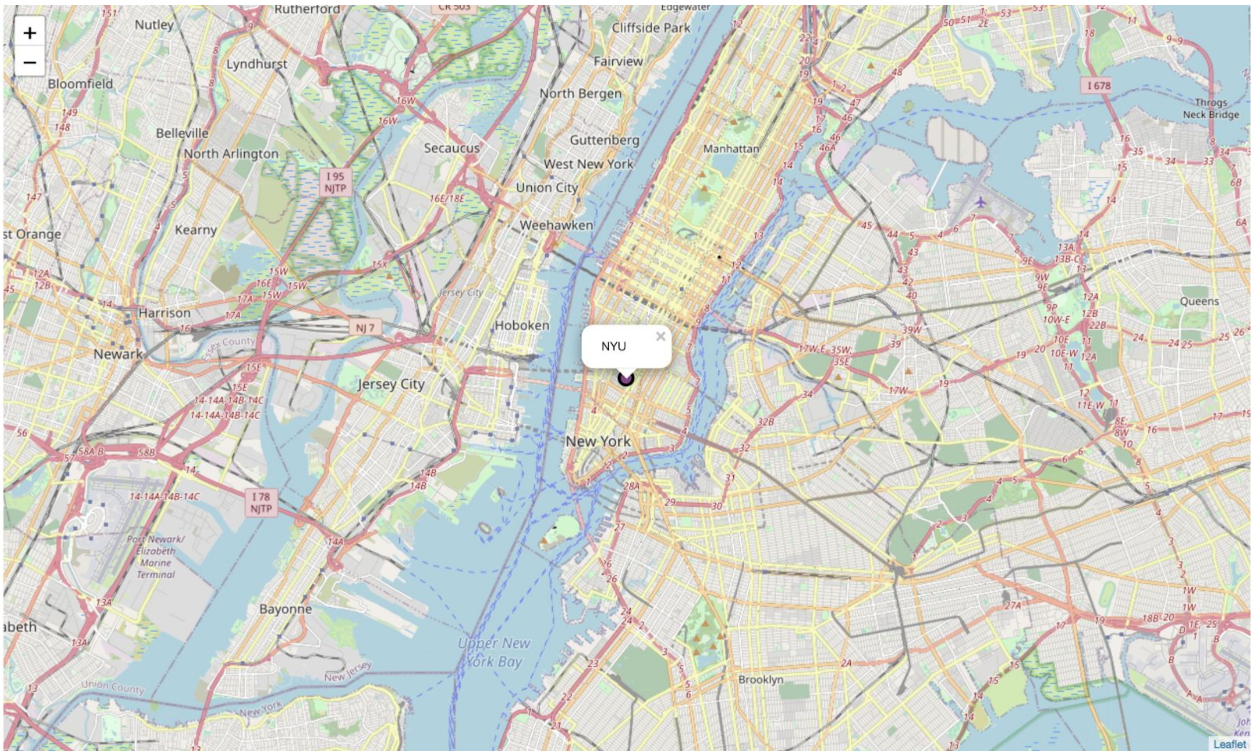
2.2 Feature Selection

We first drop features that have no value in our analysis or aren’t important for stakeholders. Three features - ‘last_review’, ‘reviews_per_month’, and ‘calculated_host_listings_count’ are dropped from our dataset. The first four features - ‘id’, ‘name’, ‘host_id’, and ‘host_name’ are all kept to ensure stakeholders can locate the listing with ease.

2.3 Data Cleaning

After dropping irrelevant features, we drop listings that are undesirable to us.

Figure 1



As shown above in Figure 1, NYU is located in Downtown Manhattan. Therefore it is reasonable for us to assume that students at NYU will generally be looking for accommodation in that area, to minimize their travel distance to and from the university. We do this by excluding listings outside of a 4 km (2.5 mi) radius of NYU.

We also set a price limit to listings, as one of the main reasons students choose to live in Airbnb’s is their price advantage over university housing. NYU housing costs around $50 per day. As Airbnb’s usually have discounts for long term stays of about 50-60%, in order for them to be significantly cheaper than NYU housing, we restrict the dataset to listings under $50.

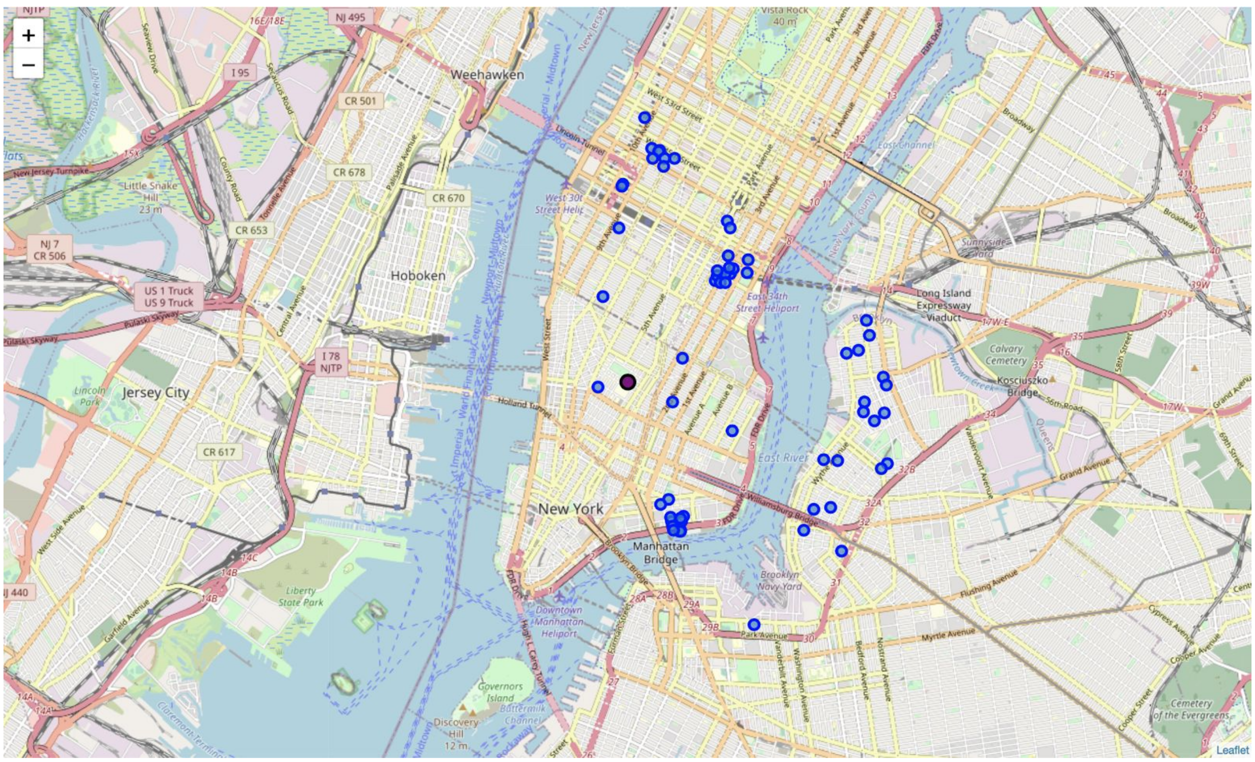
We now consider other desirable factors in listings. Reviews are fairly important to see how other people felt about the listing. But as there is no review rating in this dataset, looking at each review will have to be done manually by students. What we can do, however, is search for airbnb's with at least 1 review.

The minimum nights for a listing determines its availability to NYU students, who generally will be renting them for a semester or less (around 100 days). We therefore look for airbnb's with minimum nights under 100 days.

Finally, some listings have been set to unavailable by setting 'availability_365', the number of days in the year that the listing can be rented, to 0. Hence we eliminate these listings.

After cleaning the data, we end up with a subset of 61 Airbnb listings with characteristics desirable to NYU students. These are shown below in Figure 2. As we can observe, there is a good spread of listings around NYU, spanning both downtown Manhattan as well as parts of Brooklyn (on the other side of the East River). There are several pockets of higher density of listings, which we may explore further.

Figure 2



3. Methodology

3.1 Exploratory Data Analysis using Foursquare API

Although we have already chosen desirable listings based on their Airbnb information so far, we must also consider the types of venues around each listing. Analyzing this data and ranking the most common nearby venues to each listing can then help students choose the Airbnb they want based on the types of venues they like.

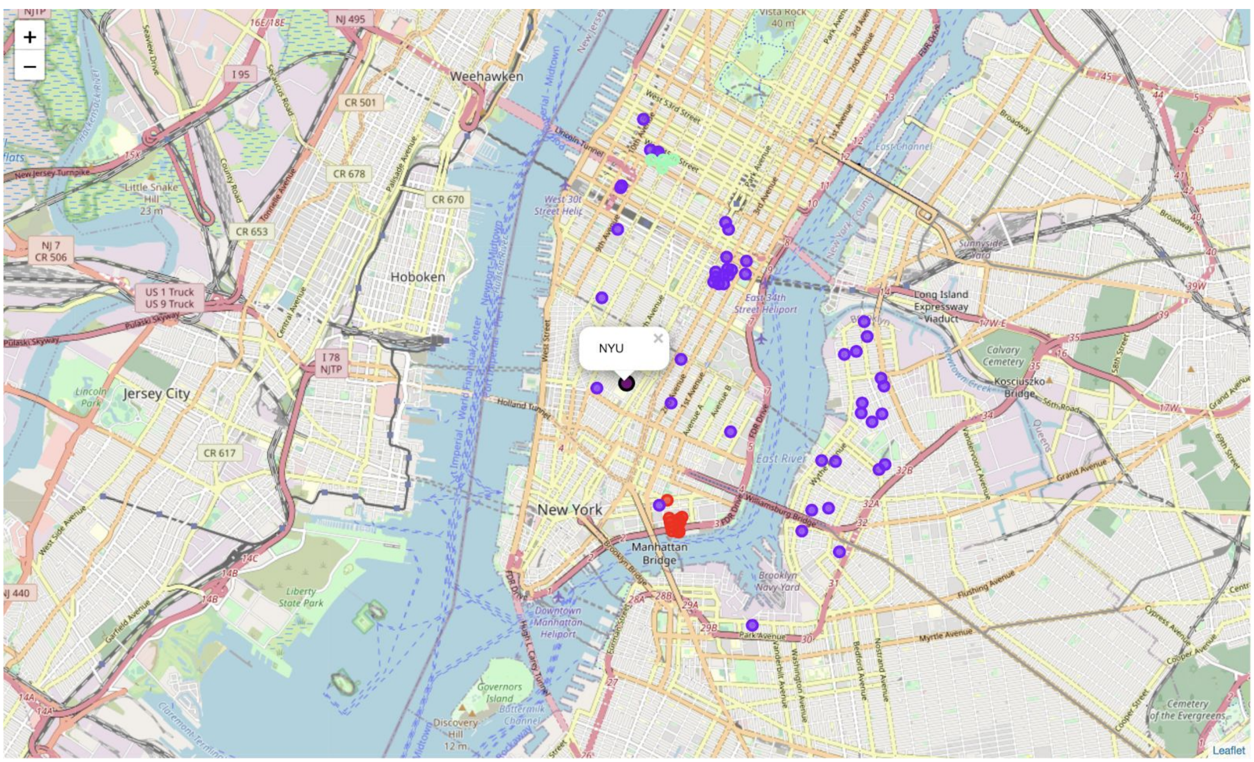
In order to find the venues nearby each listing, we utilize the Foursquare API [3]. We input coordinates of each Airbnb to the explore function to explore the kinds of venues nearby. The radius is limited to 500 meters around the listing and the number of venues is limited to 20 to reduce computation time.

To get the most common venues for each listing, we first perform one-hot encoding on the venues. We then group them by listing, as well as rank them according to the number of types of venues nearby each listing. Finally, we limit results to the top 5 to get the top 5 most common venues for each listing.

3.2 Clustering Similar Airbnbs

To gain further insight into the venues nearby each listing, we can group listings based on similar nearby venues. We use k-means clustering with 3 clusters to do this. This results in the map shown in Figure 3, with each cluster represented by a different color.

Figure 3



The first thing we may notice is the purple cluster is a lot bigger and expansive than the red or turquoise cluster, with 45 listings. Looking into it further, there are not many defining features about it, other than the large majority of venues nearby are coffee shops. Hence we name the purple cluster ‘General_Region’.

The turquoise cluster is grouped more tightly together, all predominantly located in the Hell’s Kitchen area of Manhattan. It is mainly defined by the large number of theaters around it. We therefore name this cluster ‘Theater_Region’.

Lastly, the red cluster is located in the Lower East Side of Manhattan, and from further exploration, we can see that region is the Chinatown region. The most common venues in this area confirm this; Chinese restaurants are by far the most common venues in this region. This cluster is also slightly cheaper than the other two, and we therefore name it ‘Cheaper_Chinatown_Region’.

4. Results

61 desirable Airbnb listings have been located based on the preferences of students attending New York University. These 61 listings have further been clustered into 3 clusters, which are: ‘General_Region’ with 45 listings, ‘Theater_Region’ with 4 listings, and ‘Cheaper_Chinatown_Region’ with 11 listings. An example of one of the resulting clusters dataframe, the Cheaper Chinatown Region, is shown below in Table 2. This dataframe is a combination of the original airbnb dataset, the distance metric ‘dist’ added to show distance to the NYU campus, and the 5 most common venues for each listing. ‘Cluster Labels’ represents which cluster each listing belongs to. In our case, ‘0’ is the Cheaper Chinatown Region, ‘1’ is the General Region, and ‘2’ is the Theater Region.

Figure 4

	name	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availability_365	dist	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	LowerEastSide apt share shortterm 1	Lower East Side	40.71401	-73.98917	Shared room	40	1	214	188	1.797145	0	Coffee Shop	Mediterranean Restaurant	Dumpling Restaurant	Spanish Restaurant	Mexican Restaurant
9	Shared male Room on Manhattan. Amazing view! II	Lower East Side	40.70993	-73.98700	Shared room	35	14	6	201	2.286069	0	Ice Cream Shop	Vegetarian / Vegan Restaurant	French Restaurant	Mediterranean Restaurant	Café
11	Shared male room on Manhattan with crazy view! I	Lower East Side	40.70985	-73.98724	Shared room	35	14	2	320	2.287639	0	Cocktail Bar	Vegetarian / Vegan Restaurant	French Restaurant	Mediterranean Restaurant	Café
26	Fresh and cozy male room on Manhattan III	Lower East Side	40.71110	-73.98865	Shared room	32	14	4	341	2.118048	0	Chinese Restaurant	Mexican Restaurant	Malay Restaurant	Tea Room	Cocktail Bar
30	Amazing cozy and warm male room on Manhattan IV	Lower East Side	40.71172	-73.98665	Shared room	35	14	3	322	2.111138	0	Chinese Restaurant	Coffee Shop	Mexican Restaurant	Cocktail Bar	Tea Room

5. Discussion

This project provides a good starting point for any NYU students looking for Airbnb housing. The three clusters can be treated as initial points of exploration depending on stakeholders’ preferences. Students looking for a more Chinese surrounding and cheaper living may consider the Cheaper Chinatown Region. On the other hand, students who may be more into theaters and shows may want to start looking in the

Theater Region. And if none of these seem appealing, stakeholders can begin their search within the General Region.

There are, however, many areas where this project can be refined. One major factor in considering Airbnb listings is the currentness of the data used. The data used in this project is from 2019. Since then, there could have been drastic changes to the airbnb listings from any number of random factors, such as economic changes, asset prices fluctuating, or even a global pandemic named COVID-19. Hence for findings to stay relevant, more recent data would be needed.

Another issue that could be better tackled is the number of clusters used. I chose 3 clusters through some experimentation, but in the future it may be better to perform an elbow method to determine the optimal number of clusters to use.

6. Conclusion

In conclusion, this project has picked out 61 Airbnb listings from an initial dataset of close to 50,000 based on features desirable to students attending NYU, such as price, distance to the university, and availability. These listings have been grouped based on the most common types of venues nearby, which may be good starting points of exploration for stakeholders. Future research could improve on this through using more recent data as well as using the elbow method to choose the optimum amount of clusters.

7. Citations

1. Hess, A. J. (2019, August 3). It costs $76,614 to go to NYU—but here's how much students actually pay. Retrieved from https://www.cnbc.com/2019/08/02/it-costs-76614-to-go-to-nyuheres-how-much-students-actually-pay.html
2. https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data
3. https://developer.foursquare.com