

Analyzing venues nearby the Airbnb listings in Staten Island, N.Y.

Oscar Aguilar

oe.agur@gmail.com

July 2020

Introduction

- COVID-19 has reached major fields with economic, political and social implications like tourism.
- SMEs and globally recognized companies in trouble: such as Airbnb (Gössling, 2020).
- Brian Chesky (CEO of Airbnb) expressed: “It took us 12 years to build Airbnb, and we lost almost everything in four to six weeks” (Entrepreneur, 2020).
- After the Wuhan lockdown, their booking activities showed a fall of 57.8% (Hu and Lee, 2020), laid off 25% of their employees and a debt equivalent to \$2 billion in equities (Pavlovska, 2020)
- People is not getting on airplanes (or crossing borders), there is an urge to promote local travelling, by car if required.
- Chesky suggested asking the owner of an Airbnb property: when the listing was rented, what is being done to clean it, is it being rented continuously, and local-or-regional restrictions (Krstic, 2020).
- Plus wear masks upon arrival, bring sanitizer and disinfectants, as well as an air purifier, to mention a few.

Problem

Considering New York, N.Y. as one of the most preferred and concurred cities by tourists in the world, it will be of paramount importance to analyze the venues offer of a smaller-and-near local area such as Staten Island. Using the location of Staten Island's Airbnb listings, it will be possible to find the kind of venues around them. In this way, local people near the zone from will be able to see which kind of businesses and places are around a potential Airbnb accommodation. This will benefit local tourists, Airbnb, listings' owners and local businesses.



Data Acquisition and Preprocessing

This first data set was extracted from Kaggle as a csv file. The data retrieval process started in 2008, with the aid of guests and hosts of Airbnb from N.Y. They contributed by providing information that describes the listings' generalities, availability and metrics. This continued until 2019, and three different version of the data set have been built

Feature	Type	Description
<code>Id</code>	Integer	Identification number of the listing
<code>Name</code>	Object	Name of the listing
<code>Host_id</code>	Integer	Identification number of the host
<code>Host_name</code>	Object	Name of the host
<code>Neighbourhood_group</code>	Object	Boroughs of N.Y. (e.g. Manhattan, Queens, Staten Island)
<code>Neighbourhood</code>	Object	Neighborhoods within the boroughs
<code>Latitude</code>	Float	Latitude coordinate
<code>Longitude</code>	Float	Longitude coordinate
<code>Room_type</code>	Object	Type of listing: private, shared, etc.
<code>Price</code>	Integer	Price per night
<code>Minimum_nights</code>	Integer	Minimum required number of nights to book
<code>Number_of_reviews</code>	Integer	Total number of reviews of the listing
<code>Last_review</code>	Object	Date of the last review
<code>Reviews_per_month</code>	Float	Average number of reviews the listing gets per month
<code>Calculated_host_listings</code>	Integer	Number of listings that the host owns
<code>Availability_365</code>	Integer	Number of days/year available for booking





Inspecting the type of variable of each feature, checking for missing values, and dropping some columns that were not necessary for the analysis. Features that did not contribute to the analysis were deleted, such as: id, host_id, host_name, calculated_host_listing_count and availability_365.

- The second source of information are the location-based services offered by Foursquare which consist of a RESTful service used to request JSON or XML data. Foursquare is a free location discovery app that allows users to find and share information about businesses and attractions in any part of the world. The Foursquare API allows developers to interact directly with their data platform. The different API methods allow developers to retrieve check-ins, venues, categories, tips, menus, among other data.

id	int64	id	0
name	object	name	16
host_id	int64	host_id	0
host_name	object	host_name	21
neighbourhood_group	object	neighbourhood_group	0
neighbourhood	object	neighbourhood	0
latitude	float64	latitude	0
longitude	float64	longitude	0
room_type	object	room_type	0
price	int64	price	0
minimum_nights	int64	minimum_nights	0
number_of_reviews	int64	number_of_reviews	0
last_review	object	last_review	10052
reviews_per_month	float64	reviews_per_month	10052
calculated_host_listings_count	int64	calculated_host_listings_count	0
availability_365	int64	availability_365	0
dtype: object			

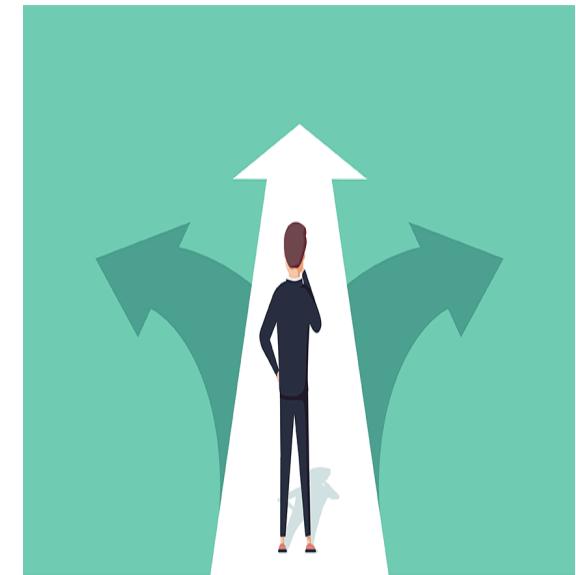
Methodology

1. From general to particular information. Techniques such as clustering, and one hot encoding are utilized. The most important libraries that were utilized are: pandas, numpy, scipy, matplotlib, seaborn, folium, geocode, and Sklearn.
2. Broad inspection of the data was made. Comparisons between the types of rooms and their mean price was performed. Similarly, the mean price of the listings per borough was obtained.
3. Staten Island listings were kept (local area), from there, 4 different maps were constructed: one with all the listings, and one per type of room (entire homes/apartments, private rooms, and shared rooms). After that, the listings were grouped by neighborhood, and the analysis showed that the neighborhood with more listings was St. George with 48.

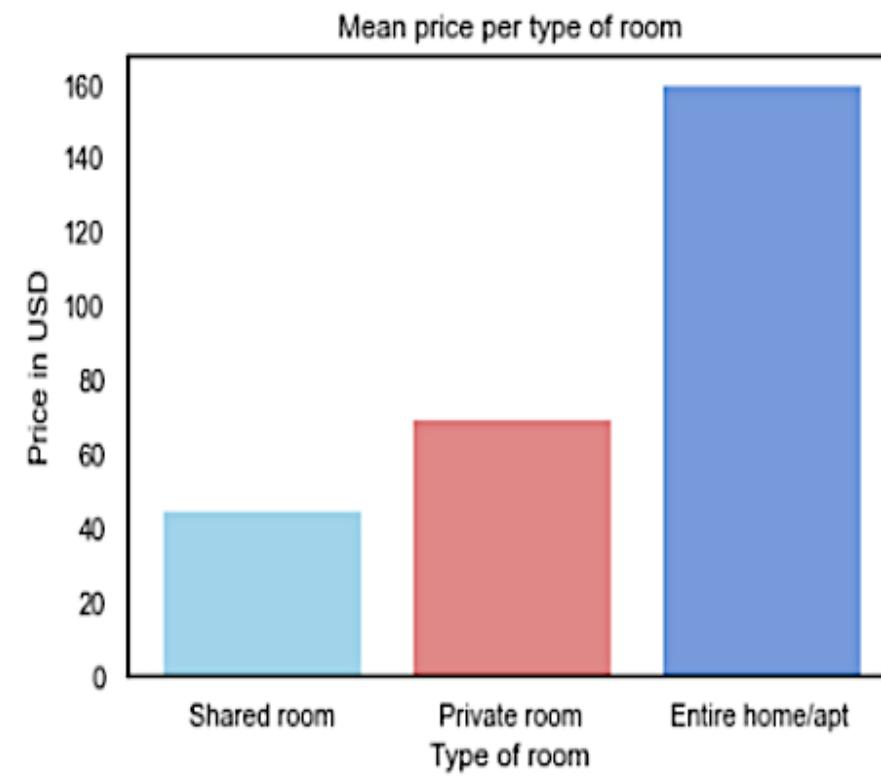
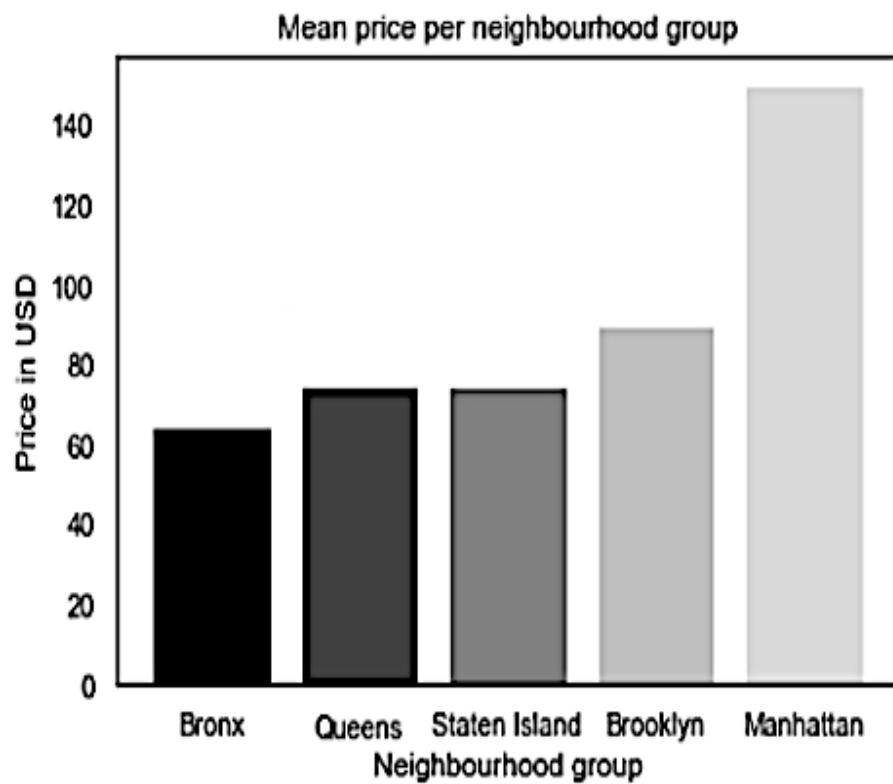


Methodology

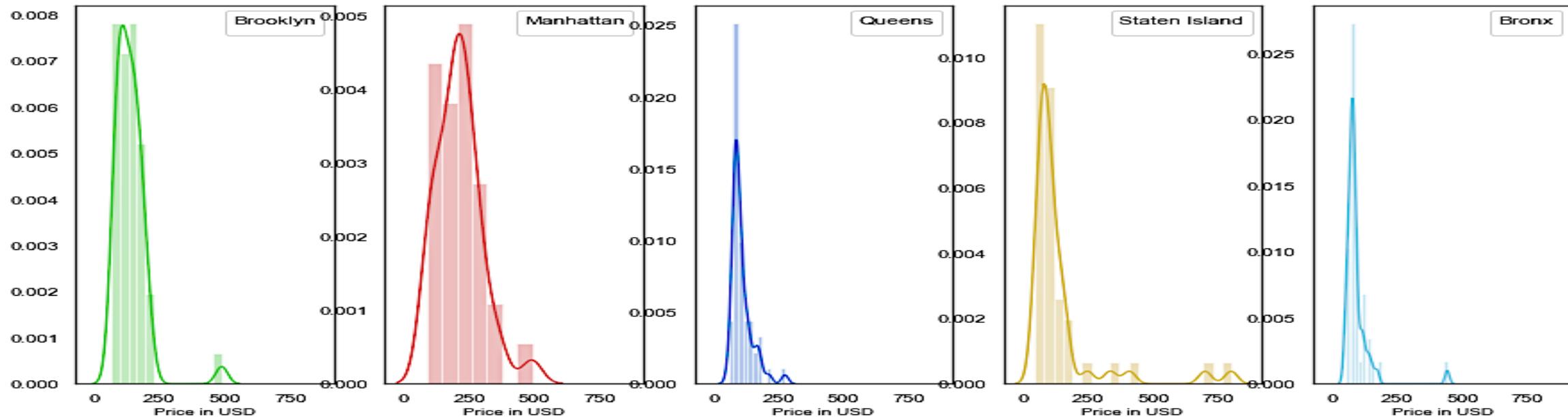
4. From there, the Foursquare API was used to retrieve JSON data from the venues of all the neighborhoods. The retrieved information showed that in Staten Island there were around 5,663 venues with 219 unique categories.
5. Next, the top 10 most common venues per neighborhood were found and displayed. This allowed to retrieve the most common venue per neighborhood and construct a graph that shows the category and count of the most common venues in the island.
6. For the interest of the Airbnb hosts, a clustering algorithm was used to group all the listings of the island in 6 different clusters. After that, the mean price of each cluster was calculated. Similarly, the clusters were visually represented in a map of the island.



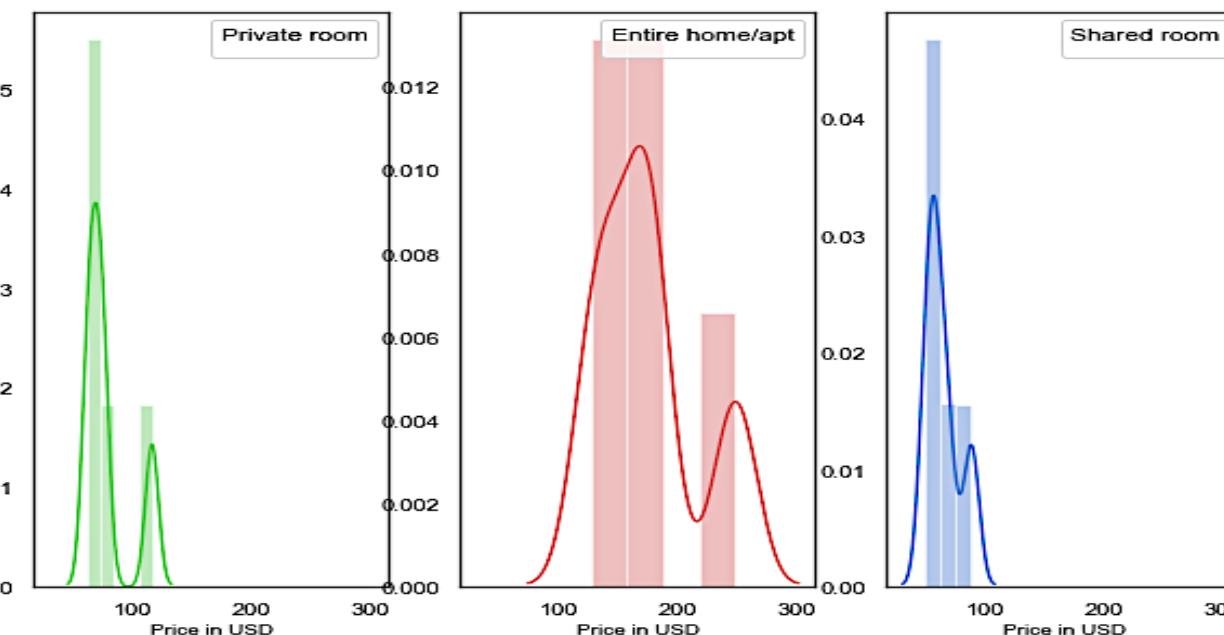
Results – Broad NYC Analysis

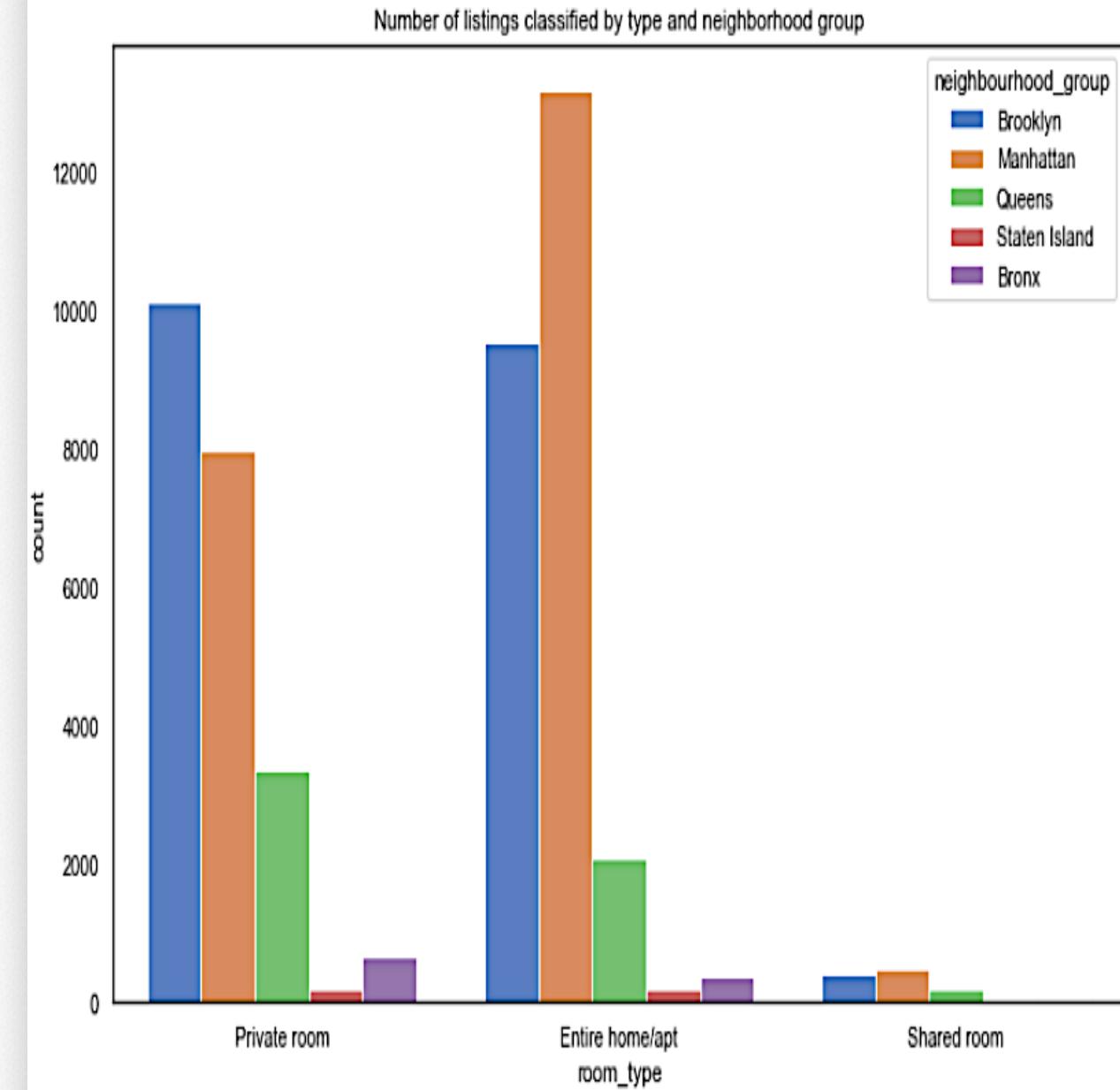
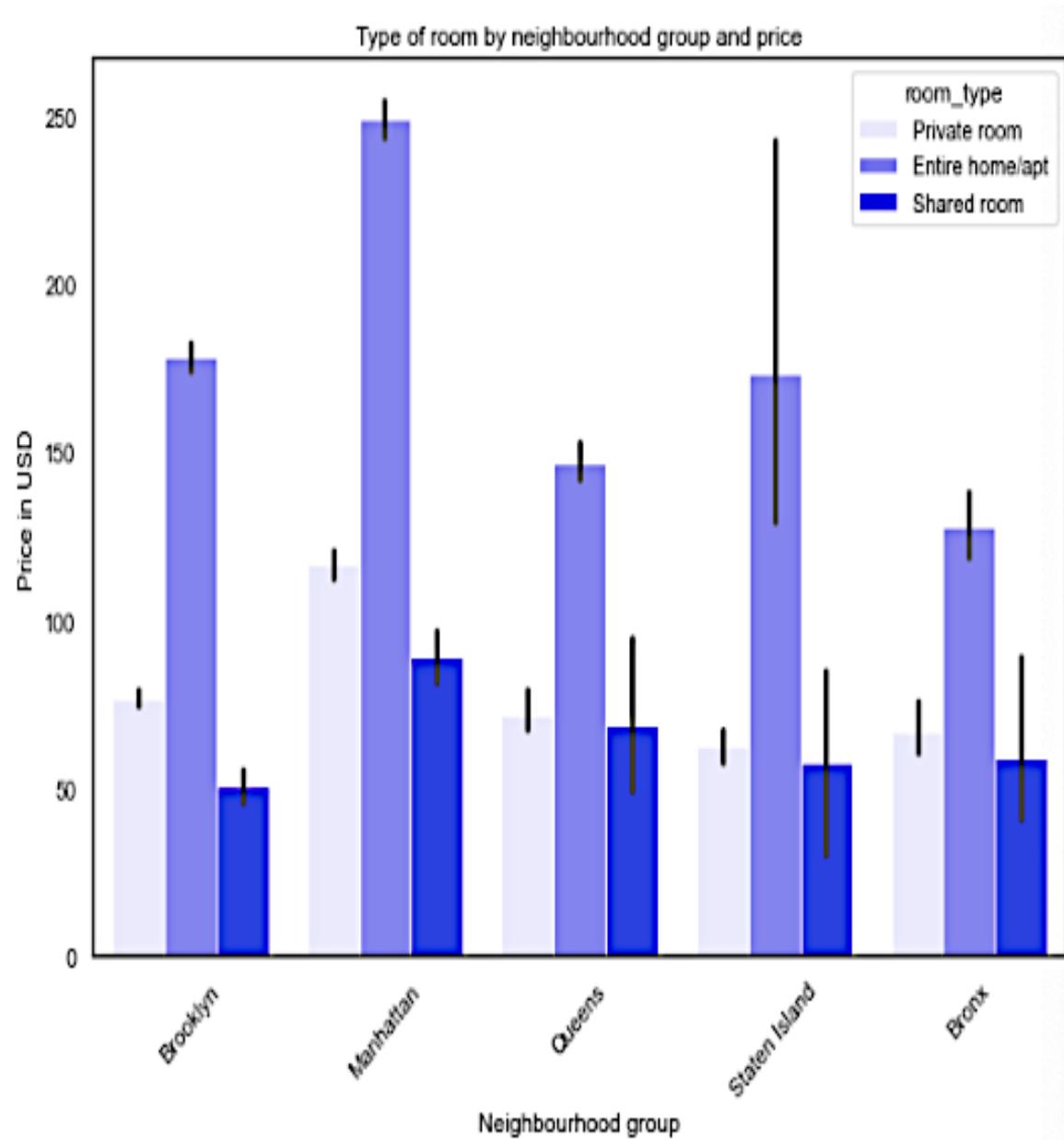


Price distribution per neighbourhood group



Price distribution per type of room





Results – Staten Island Analysis

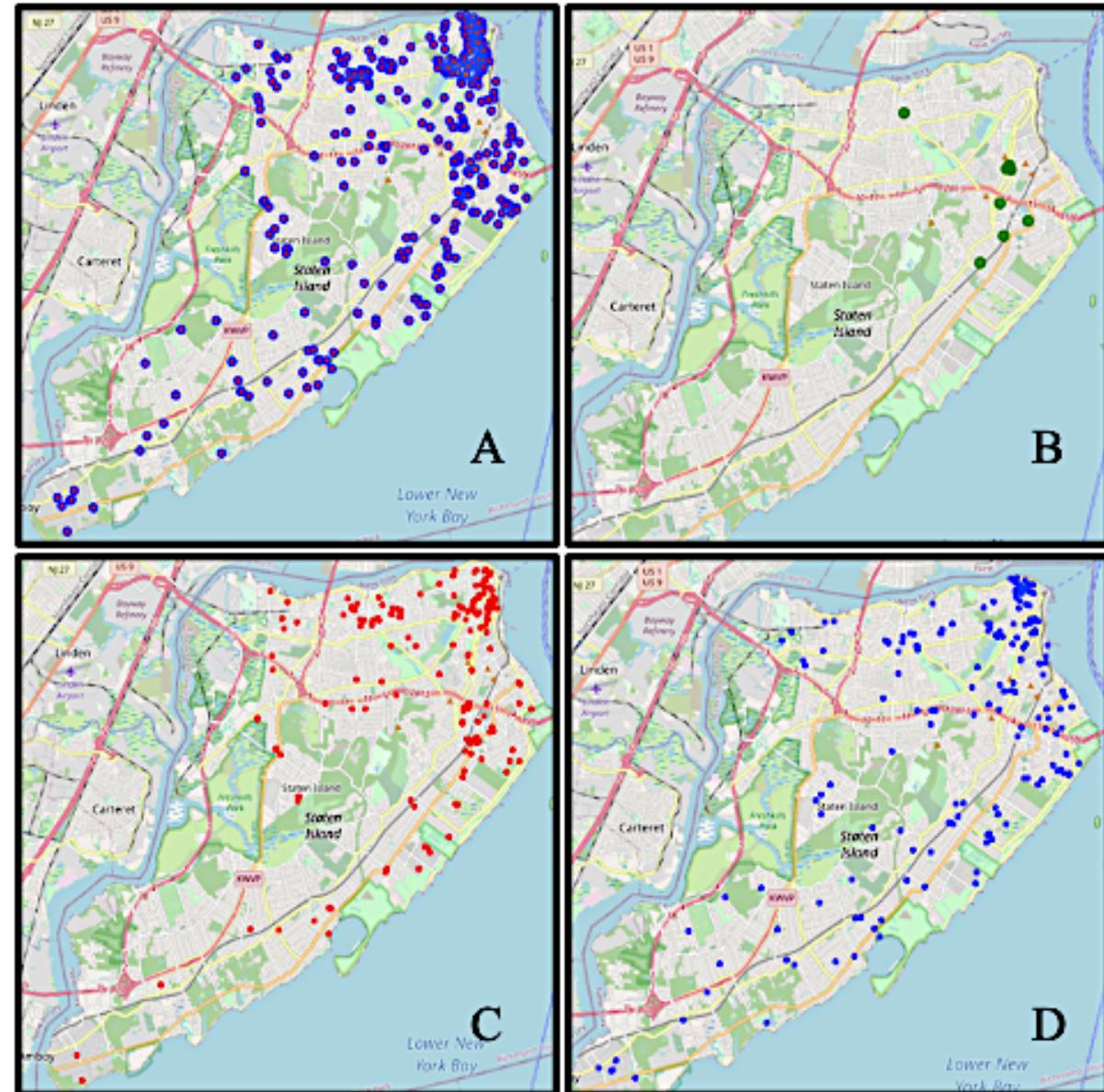
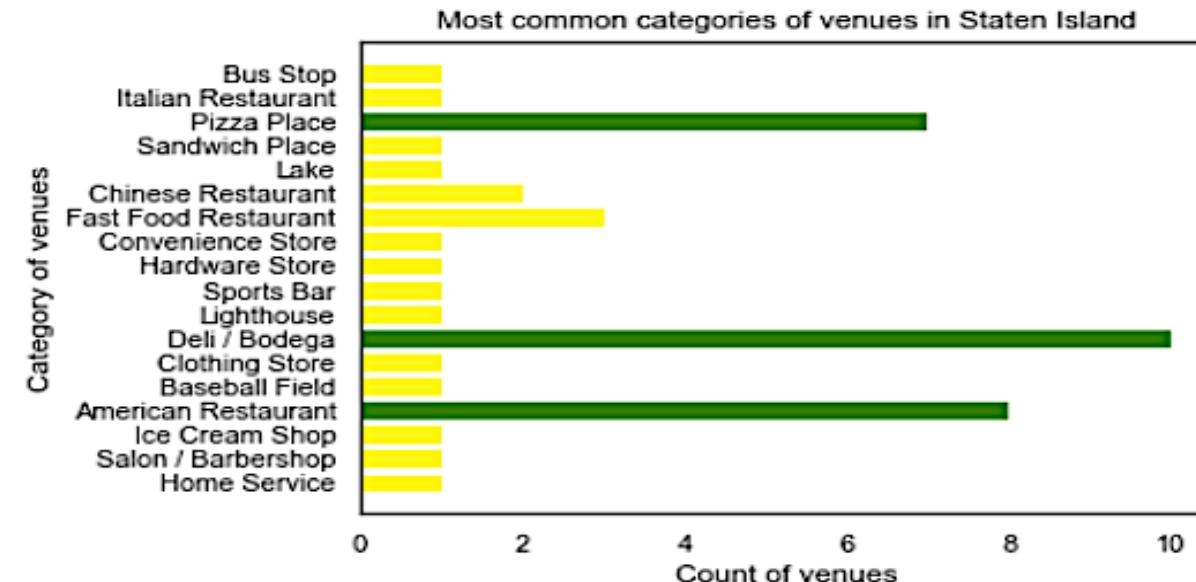
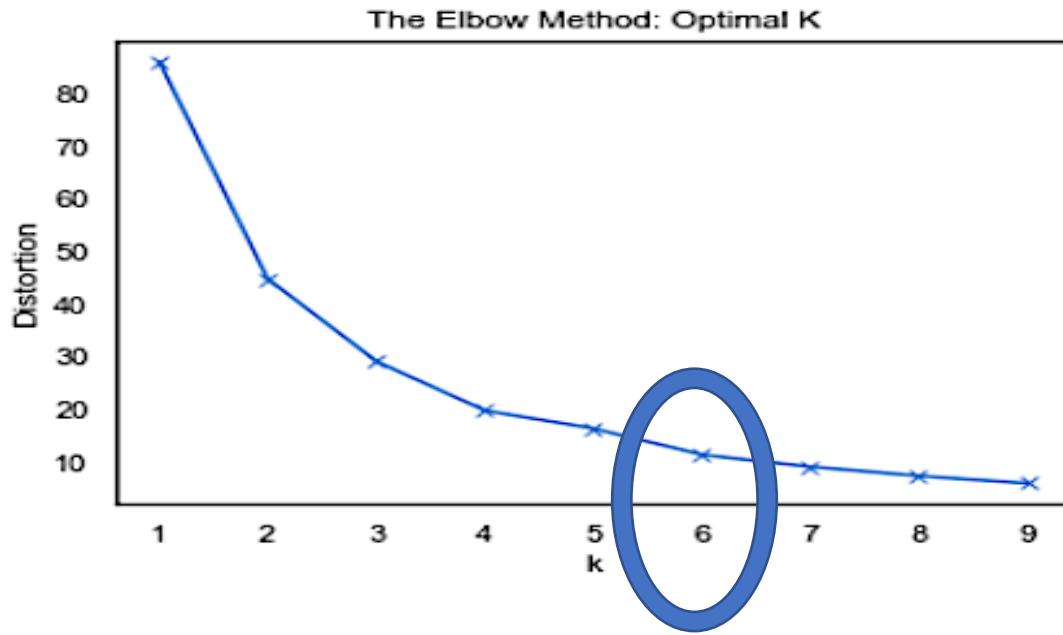


Figure 8. A) Geographical location of listings B) Shared rooms C) Private rooms D) Entire homes/apartments

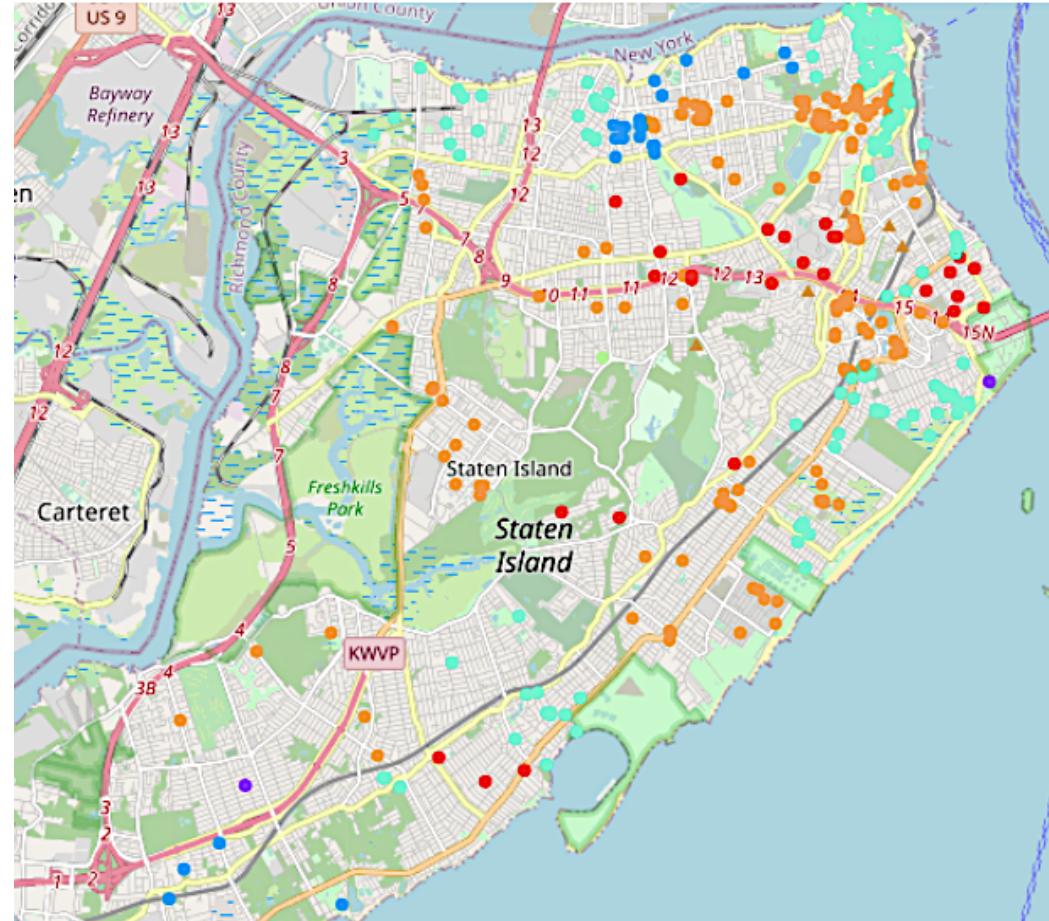
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arden Heights	Bus Stop	Asian Restaurant	Park	Diner	Chinese Restaurant	Restaurant	Pizza Place	Pub	Pool	Liquor Store
1	Arrochar	Italian Restaurant	Pharmacy	Baseball Field	Liquor Store	Deli / Bodega	Bakery	Beach	Pizza Place	Sandwich Place	Grocery Store
2	Bay Terrace, Staten Island	Italian Restaurant	Insurance Office	Supermarket	Plaza	Bar	Pizza Place	Playground	Chinese Restaurant	Donut Shop	Salon / Barbershop
3	Bull's Head	Pizza Place	Health & Beauty Service	Coffee Shop	Chinese Restaurant	Food	Bagel Shop	Spa	Sandwich Place	Baseball Field	Pharmacy
4	Castleton Corners	Pizza Place	Bank	Chinese Restaurant	Ice Cream Shop	Diner	Optical Shop	Mini Golf	Sandwich Place	Bus Stop	Flower Shop



Results – Clustering Analysis



Cluster	Mean price in USD
#1	152.12
#2	750.00
#3	348.78
#4	106.72
#5	249.00
#6	71.55



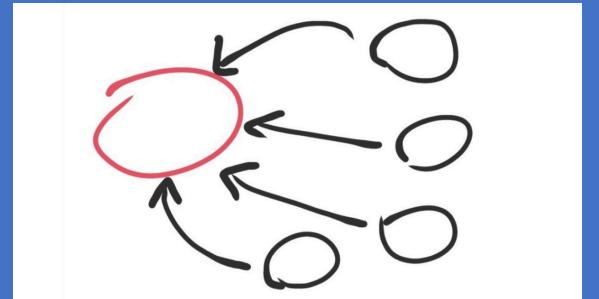
Discussion



- Near New York City, there are some local areas that have a lot to offer to people near the surroundings.
- In the case of Staten Island, there plenty of restaurants to go to (especially pizza places and American food). On the other hand, there are outdoor activities to do such as baseball, lake visiting, lighthouse exploring, among others. Nonetheless, the top venues were pizza places, delis/bodegas, and American restaurants.
- On the other hand, all the listings within Staten Island were clustered into 6 different groups. With this information, the mean price of the listings of each clustered was computed. This information will allow the owners of the listings to analyze if their prices are competitive or not. The latter will also help them to plan accordingly to the new normal era of tourism, which is more local and in shorter periods.
- The whole data set was analyzed too. The main findings were not surprising, due to the fact that showed that the most expensive Airbnbs were situated within Manhattan. It also showed that the most expensive type of listings were entire homes and apartments. The price distribution of each type of listing by borough was graphed and showed that Staten Island offered listings with a wide range of prices, which also inspired the project to take the analysis further.

- While the pandemic of COVID-19 remains among us, the tourism sector will need to readapt their clauses in order to survive. In term of accommodation, some actions will need to be taken. However, one of the most common aspects of this new normality, is that tourism needs to be more local. Tourists will be forced to travel only by car, and the most obvious thought is that they are going to visit near locations.
- In this project, under the assumptions mentioned above, a local area near the most important city of the world was analyzed. The Airbnb listings within Staten Island were retrieved and positioned in the map; furthermore, the most common venues were obtained and studied. Similarly, these listings were clustered into 6 different groups and their mean price was computed. The latter will allow the owners of the listings to reassure or restructure their prices, in order to be more competitive.

Conclusion



References

- Entrepreneur staff (2020). Airbnb CEO: It Took Us 12 Years to Build, and We Lost Almost Everything in 6 Weeks. *Entrepreneur magazine*
- Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of COVID-19. *Journal of Sustainable Tourism*, 1-20.
- Hu, M. R., & Lee, A. D. (2020). Airbnb, COVID-19 Risk and Lockdowns: Local and Global Evidence
- Krstic, Z. (2020). Is it safe to Rent an Airbnb? How to Lower COVID-19 Risks in Vacation Homes
- Pavlovska, E. (2020). “Airbnb CEO says post-COVID travel will never be the same again” in *New Europe*.