Trending Seattle Venues and Near Airbnb Data Analysis

Wilton Beltré
Septiembre 28, 2019

1. Introduction

1.1 Background

Boasting a population of over 724,000, Seattle is one of America's most scenic cities. Home to major tech corporations such as Microsoft, Tableau, and Amazon, Seattle is also known for its breathtaking mountain ranges, national parks, and lush green scenery. Given this balance between a robust tech scene and a beautiful landscape, it is no wonder that Seattle is a popular destination for major conferences and events.

1.2 Interest

There are many destinations in the Pacific Northwest that are culturally interesting, full of excellent places to eat and wine, and have a good positive vibe about them.

So I will show you an observatory study that, in a tactical way, can be very useful for real estate agencies as well as for the individual who is looking for attractions in Seattle and a place to stay, as it can be very useful to measure the vibe of the neighbors and answers the following questions:

- Where is the best place to open a retail business with a good start?
- Which neighborhood is similar to another neighborhood in Seattle?
- What are the most trending neighborhoods according to their attractions?
- Where in Seattle should I rent an apartment?

With my study I am trying to answer these and other questions that will be very useful for both now and the future.

I hope it is of good profit for all the stakeholders.

2. Data Acquisition

I have considered supporting the solution with the following data sources listed below

- The list of Seattle neighborhoods was found in Wikipedia but without coordinates, so I made a procedure to find all coordinates using geopy Python libraries.
  https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle

- I used Foursquare API to get the *trending* venues of given Borough of Seattle.

- Looking at kaggle.com I could find the scrapped well done data of listings of current Airbnb in Seattle. This dataset is hosted at https://www.kaggle.com/shanelev/seattle-airbnb-listings and was created with the help of Tom Slee's Airbnb, Data Collection codebase that can be found at https://github.com/tomslee/Airbnb-data-collection

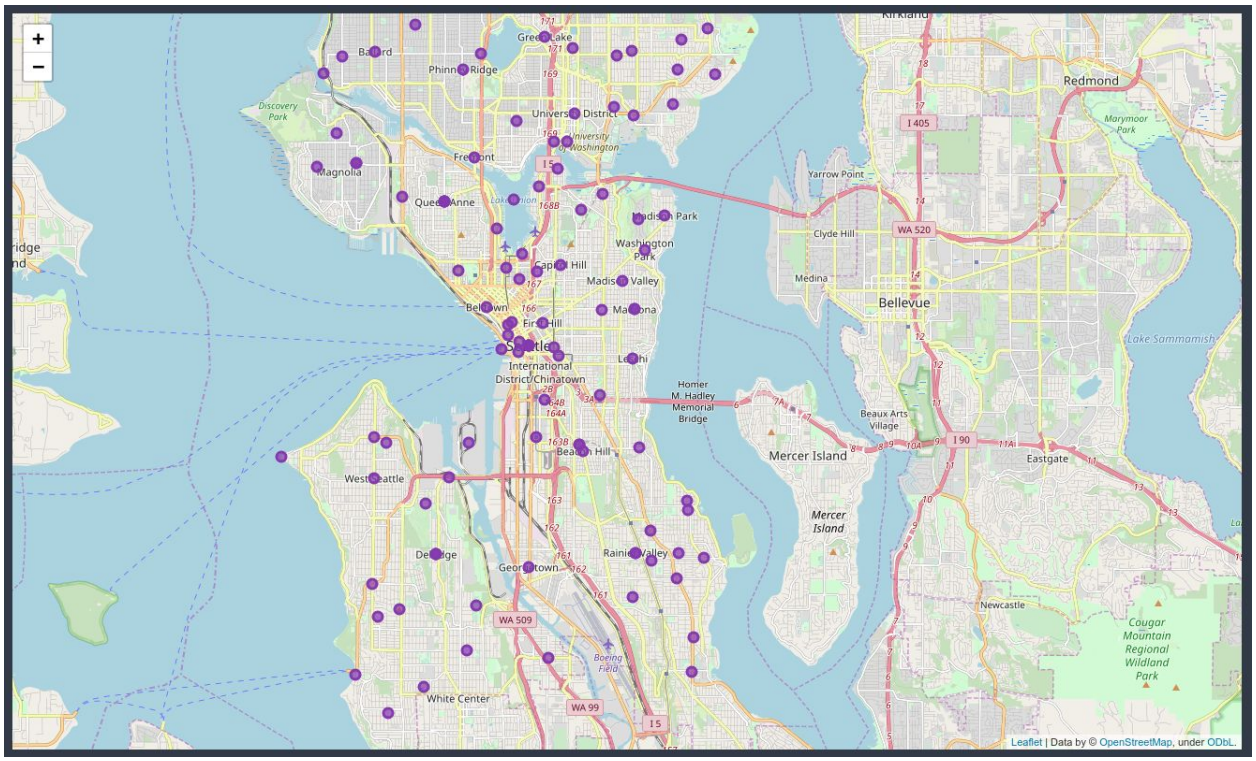
- To get a dataset that has the venues and its relationship with the closest Airbnb, I did the following:
    - Searched for each venue: the closest Airbnb to this venue.
    - Used the geopy support that allows measuring the distance from point to point, to determine who is less than 200 meters away.
    - For convenience, the result of previous process was saved to file due to the large time of processing.
    - Thanks to the support of the pandarallel library (https://github.com/nalepae/pandarallel) processing could be improved by doing 8 separate tasks to each core of the machine processor.

3. Methodology

As I mentioned before, my data are from different sources, and they were combined to moderate the expected result. My master data has the main components: Borough, Neighborhoods, Trending Venues with Latitude and Longitude and Airbnb rent data with quantity of near to venue and Latitude and Longitude respectively.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	airbnb_near	distance_km	near_airbnb_latitude	near_airbnb_longitude
0	North Seattle	47.660773	-122.291497	QFC	47.662216	-122.296431	Supermarket	0	NaN	NaN	NaN
1	North Seattle	47.660773	-122.291497	China Village	47.661396	-122.291437	Chinese Restaurant	2	[0.1751971902,0.1528751892]	[47.662065,47.660591]	[-122.289325,-122.289787]
2	North Seattle	47.660773	-122.291497	T-Mobile	47.661474	-122.296080	Mobile Phone Shop	0	NaN	NaN	NaN
3	North Seattle	47.660773	-122.291497	Burke-Gilman Brewing Company	47.661308	-122.288067	Brewery	2	[0.1265926103,0.1518224476]	[47.662065,47.660591]	[-122.289325,-122.289787]
4	North Seattle	47.660773	-122.291497	General Porpoise Coffee & Donuts	47.662010	-122.293218	Donut Shop	0	NaN	NaN	NaN

I used python folium library to visualize geographic details of Seattle and its boroughs and I created a map of Seattle with boroughs superimposed on top. I used latitude and longitude values to get the visual as you can see below:



I utilized the geopy Python API to acquire the boroughs' coordinates. Using Foursquare to get venues, I designed the limit as 100 venue per neighborhood and the radius of 500 meter for each borough from their given latitude and longitude information.

DataFrame scrapped from Wikipedia and helped with geopy

	Borough	Neighborhood	latitude	longitude
0	Seattle	North Seattle	47.660773	-122.291497
1	North Seattle	Broadview	47.722320	-122.360407
2	North Seattle	Bitter Lake	47.726236	-122.348764
3	North Seattle	North Beach	47.696210	-122.392362
4	North Seattle	Crown Hill	47.694715	-122.371459

DataFrame combined from Foursquare

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	North Seattle	47.660773	-122.291497	QFC	47.662216	-122.296431	Supermarket
1	North Seattle	47.660773	-122.291497	China Village	47.661396	-122.291437	Chinese Restaurant
2	North Seattle	47.660773	-122.291497	T-Mobile	47.661474	-122.296080	Mobile Phone Shop
3	North Seattle	47.660773	-122.291497	Burke-Gilman Brewing Company	47.661308	-122.288067	Brewery
4	North Seattle	47.660773	-122.291497	General Porpoise Coffee & Donuts	47.662010	-122.293218	Donut Shop

3.1 Airbnb Seattle Dataset

The data was *scrapped* on December 19th, 2018 and contains roughly 8000 listings of current Airbnb listings in Seattle. The data has the price, reviews, latitude, longitude, bedroom, bathroom, number of guests it accommodates, room type, and more. For this purpose I only needed latitude and longitude as well.

Airbnb dataset of Seattle

	latitude	longitude	room_type	name
0	47.610819	-122.290816	Entire home/apt	Casa Madrona - Urban Oasis, 1 block from the P...
1	47.529846	-122.275840	Entire home/apt	Sweet Seattle Urban Homestead 2 Bdr
2	47.687801	-122.313427	Private room	Sunrise in Seattle Master Suite
3	47.523980	-122.359891	Entire home/apt	Cozy Studio, min. to downtown -WiFi
4	47.654109	-122.337605	Entire home/apt	Fab, private seattle urban cottage!

3.2 Heavy Work: Combining Seattle Dataset

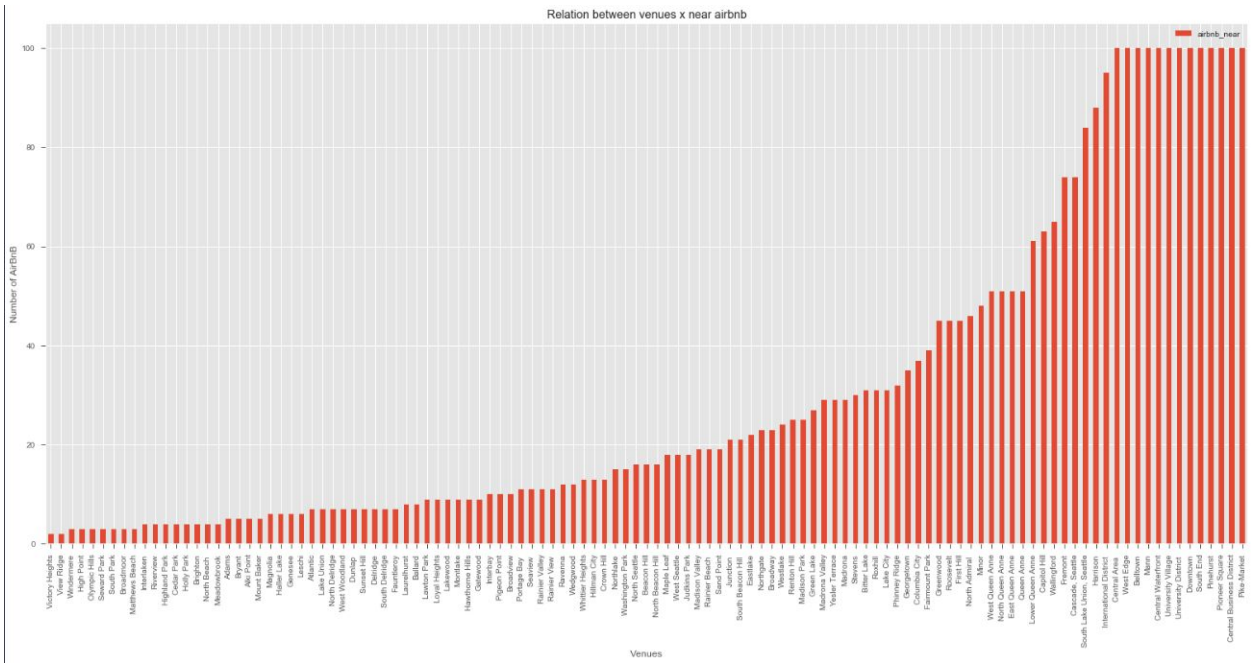
I was working with my algorithm to find each property for rent with a proximity of 200 meters or less from any given venue.

At the end, this DataFrame figures out as follows, where you may see the venue's latitude and longitude with an array of respectively near Airbnb coordinates and distance from these.

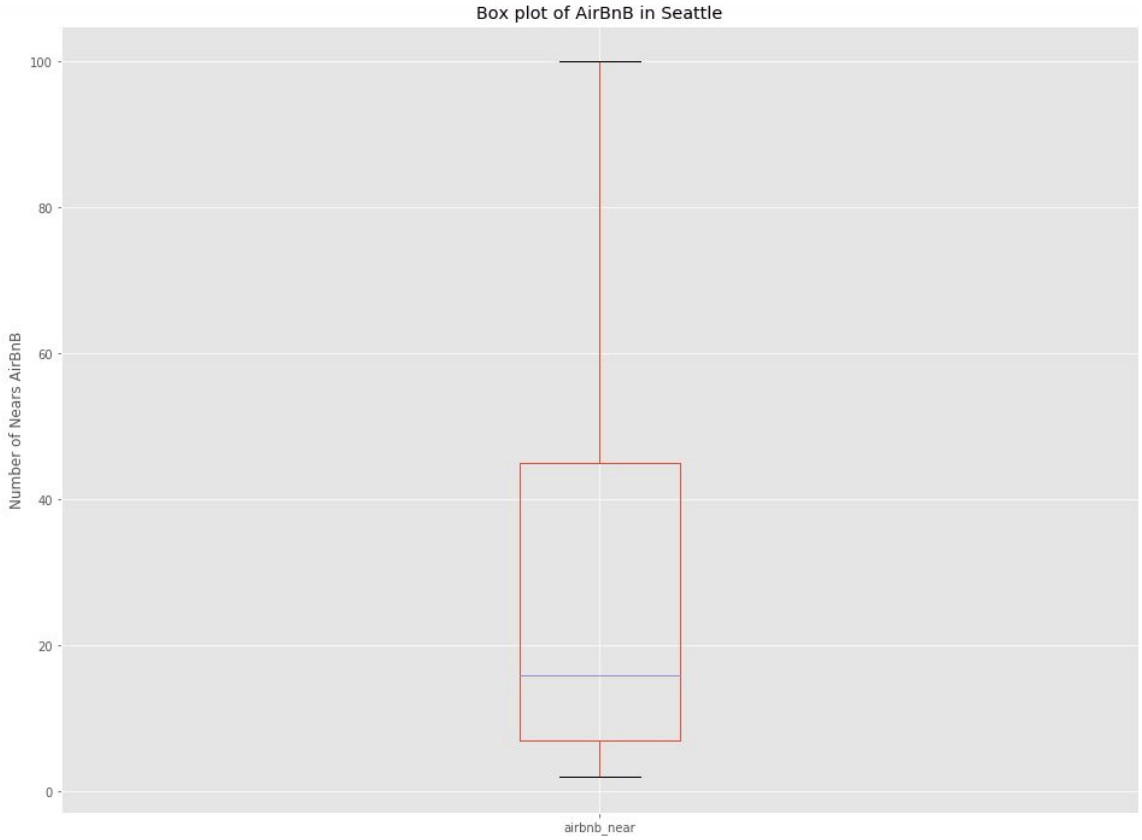
		distance_km	near_airbnb_latitude	near_airbnb_longitude
venue_latitude	venue_longitude			
47.516686	-122.366630	[0.1525084617,0.1594414796]	[47.515423,47.516495]	[-122.36584,-122.364532]
47.516689	-122.366683	[0.1544129672,0.1634633405]	[47.515423,47.516495]	[-122.36584,-122.364532]
47.516797	-122.369529	[0.1183192174,0.135584232]	[47.515741,47.515764]	[-122.36934,-122.370485]
47.517139	-122.368917	[0.1587004585,0.1932021458]	[47.515741,47.515764]	[-122.36934,-122.370485]
47.519691	-122.267145	[0.1544736713]	[47.51929]	[-122.269109]

3.4 Statistical information of the Airbnb near the venues

Using grouping technique, first proceed to count the Airbnbs of each venue and then sort it in ascending for easily visualization in a **Plot Bar**.

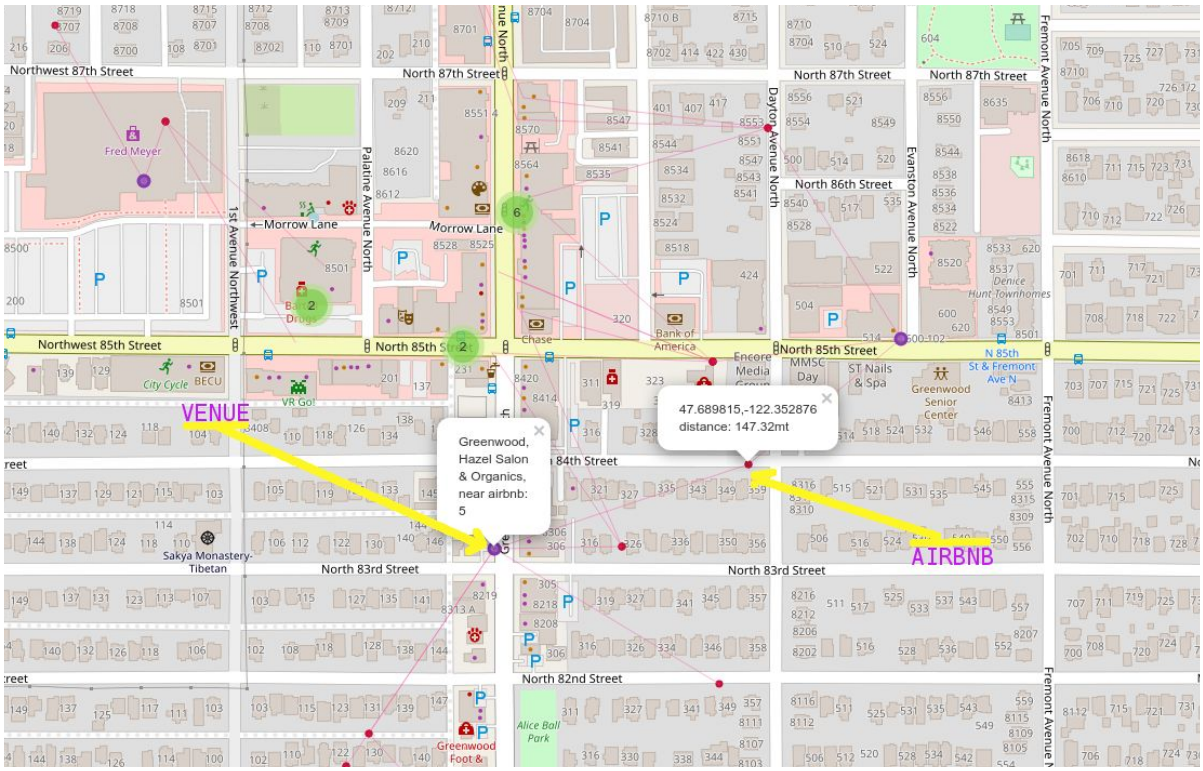


Box Plot allows you to quickly appreciate the average, minimum and maximum as other information about the relationship we have been observing.



3.5 Showing on Map the trendings venues of Seattle and their connection with near Airbnb

This is the result obtained with the proximity algorithm. To build a DataFrame like this one allows us to see the beauty of knowing that the red dots are the Airbnb rental locations drawn with a line that shows the distance between a place that are mutually connected.



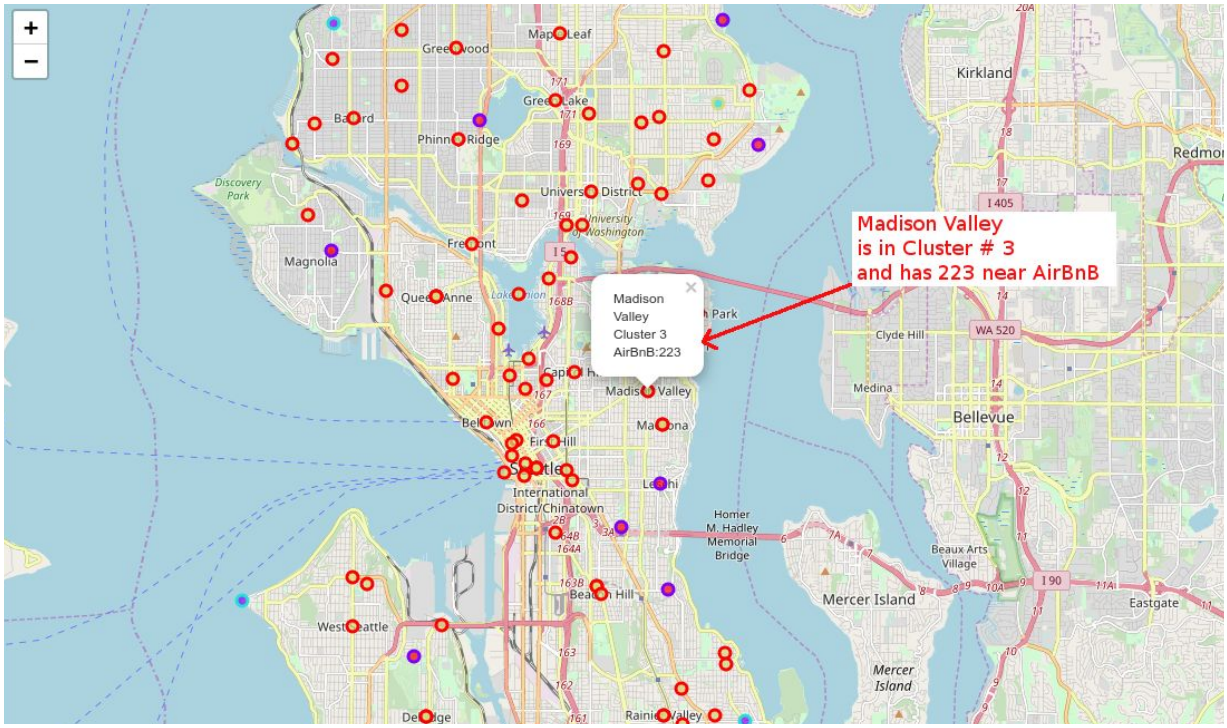
3.6 Cluster Neighborhoods

First of all we prepare the data by grouping the neighborhoods and count the fields outside the grouping. To continue with generating of dummies, which is a Pandas library facility, in order to transpose the data to initially measure the amount of venues per neighborhood.

Now that the DataFrame is grouped by neighborhood, let's look for the average of each venue category, this is very simple when applying `GroupBy.mean()` function.

I still need an indicator with the amount of properties for rent per venue, so that it can be compatible, it must be grouped equally by neighborhood and sum up the amount of properties for rent from Airbnb.

The map is ready to be observed, the Clusters can be fed with the previous merged DataFrame and adapt it for analysis.



4. Conclusion

In a single view the segmentation by groups or clusterization has been taken into consideration, the identification of groups gives us many answers to the main questions and in addition to this the Airbnb vibe is available for each Neighborhood.

	Neighborhood	airbnb_near	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
14	Cedar Park	0	Lake City	47.725653	-122.285960	0	Vietnamese Restaurant	Supermarket	Fast Food Restaurant
29	Gatewood	24	West Seattle	47.536560	-122.385326	0	Park	IT Services	Pet Store
30	Genesee	22	West Seattle	47.564821	-122.367905	0	Gym	Park	Golf Course
34	Haller Lake	4	Northgate	47.720781	-122.330424	0	Food Truck	Playground	Coffee Shop
37	High Point	3	Delridge	47.538451	-122.377387	0	Park	Field	Playground
44	Judkins Park	76	Atlantic	47.591470	-122.304069	0	Park	Museum	Rental Service
51	Leschi	32	Central Area	47.600535	-122.291866	0	Park	Grocery Store	Pet Store
58	Magnolia	15	Central Seattle	47.648937	-122.393115	0	Park	Dog Run	Gym / Fitness Center
61	Matthews Beach	1	Lake City	47.696927	-122.272724	0	Park	Trail	Playground
65	Mount Baker	4	Rainier Valley	47.578552	-122.289526	0	Park	Thai Restaurant	Pet Store

References

- *https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle*
- *https://github.com/tomslee/Airbnb-data-collection*
- *https://github.com/nalepae/pandarallel*
- *https://www.kaggle.com/shanelev/seattle-airbnb-listings*
- *https://github.com/tomslee/airbnb-data-collection*
- *https://foursquare.com/developers/explore*