

Where do I move into in Seattle?

Introduction to the Problem

I recently visited Seattle, and I enjoyed my time there. I was researching what it would take to move there, and I came across the idea for my Capstone project.

For people moving to Seattle, how can we find a good neighborhood to move into?

This question can be answered in a myriad of ways, but I am trying to answer it by clustering 12 of the best neighborhoods in Seattle based on the frequency of different venues at their locations.

I will try to find neighborhoods that have a unique list of venues and are affordable compared to other neighborhoods that are considered. Using k-means clustering and increasing the number of clusters, I can find neighborhoods that have similar kinds of venues in the Foursquare database.

Project Audience

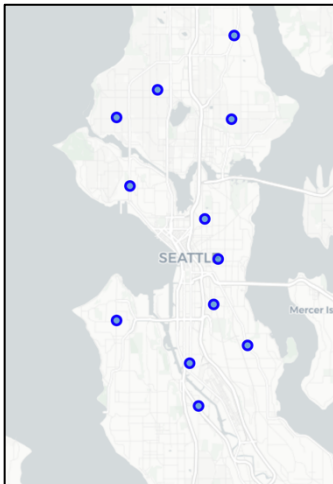
This project will help people who are deciding to move to Seattle, who will be the target audience. It will help them decide a neighborhood to look for apartments in, at an affordable rate.

Expected Result

Usually there is wiggle room with respect to the amount of rent a person is willing to pay, so the result of this project is to get a visual representation of the data that helps make an informed choice.

Data

For this project, I will consider 12 neighborhoods, the average rent at these places, and the uniqueness of venues in a 500-meter radius of each neighborhood's location.



The image to the left is a map of Seattle with the blue markers representing the neighborhoods that are considered. They are listed as follows: West Seattle, Beacon Hill, Ravenna, Columbia City, Ballard, Interbay, South Park, Central District, Georgetown, Lake City, Greenwood, and Capitol Hill.

Neighborhood name and the Average Rent in USD are obtained by web scraping the following website:
<https://www.seattlemet.com/home-and-real-estate/2019/03/the-top-12-neighborhoods-in-seattle-2019-edition>

A sample of the website that is web-scraped is shown in the screenshot below:

Beacon Hill

Median Sale Price: **\$585,750** • Median Sale Price YOY: **2%** • Homes Sold in 2018: **80**

• Average Rent: **\$1,634** • Walk Score: **63** • Transit Score: **56**

After web scraping, the data is stored as a .csv file, a screenshot of which is shown below:

	A	B	C	D	E
1		Neighborhood	Average Rent (USD)	Latitude	Longitude
2	0	West Seattle	1832	47.5709315	-122.38652
3	1	Beacon Hill	1634	47.579271	-122.31179
4	2	Ravenna	1598	47.6756537	-122.29763
5	3	Columbia City	2272	47.5579124	-122.28522
6	4	Ballard	2174	47.6765073	-122.38622

Data obtained using the Foursquare API will also be used for the project. Foursquare data will be queried using the API to get a list of venues, and the categories in which those venues lie.

A sample of the data in .csv format is shown below:

	A	B	C	D	E	F	G	H
1		Neighborhood	Neighborhood	Neighborhood	Venue	Venue Latitude	Venue Longitude	Venue Category
2	0	West Seattle	47.5709315	-122.38652	Olympia Coffee	47.5693838	-122.38668	Coffee Shop
3	1	West Seattle	47.5709315	-122.38652	Meeples Games	47.5704955	-122.38704	Gaming Cafe
4	2	West Seattle	47.5709315	-122.38652	Prost - West Seattle	47.5734936	-122.38696	Pub
5	3	West Seattle	47.5709315	-122.38652	West Seattle Runner	47.5703625	-122.38709	Sporting Goods Shop
6	4	West Seattle	47.5709315	-122.38652	Spiro's Pizza & Pasta - West Seattle	47.573744	-122.38696	Pizza Place

After web scraping for neighborhoods data, I used *Nominatim* to geocode the neighborhood locations to get their longitude and latitude. I then accessed location data for venues using the Foursquare API.

I then encoded the venue categories for each neighborhood and created a *pandas* dataframe with weighted frequency of venues for each neighborhood.

This completed the data acquisition and cleaning part of the project.

Methodology

To find neighborhoods that have similar venues, we use a K-means clustering algorithm. Clustering measures the similarity of factors (e.g., venue frequency) to cluster neighborhoods together. Since we are only considering 12 neighborhoods, most neighborhoods that are similar

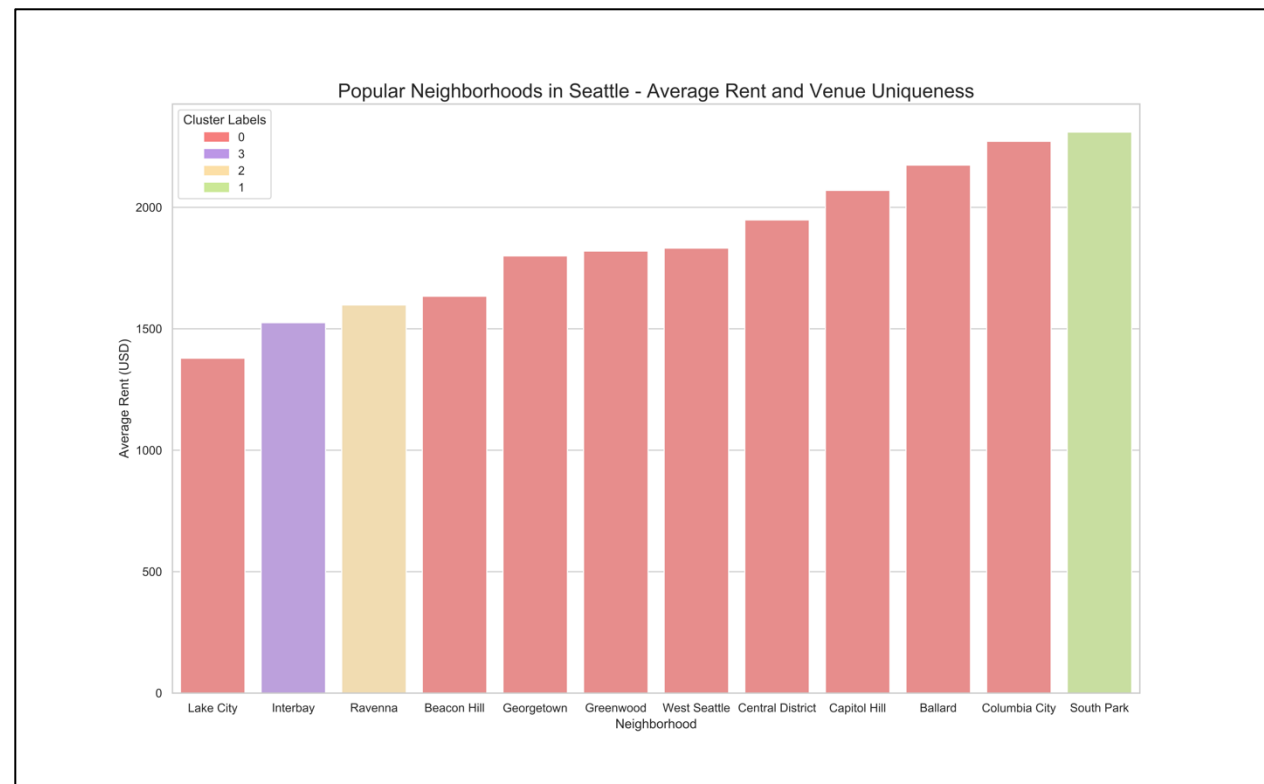
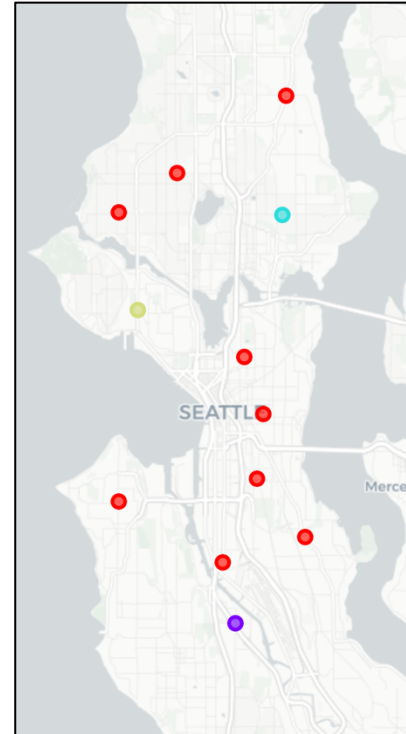
to each other will be clustered together, whereas unique neighborhoods will be a part of their individual clusters.

I started by sorting the neighborhoods into 3 clusters first, and then I moved on to sorting neighborhoods into 4 clusters to see if any bigger clusters were formed. Sorting these neighborhoods into even more clusters overfits the data, as we would see neighborhoods that are similar to be split into different clusters, which we do not want.

Results on the map of Seattle after 4 clusters are shown to the right. The red markers represent neighborhoods that were clustered together into the first cluster (cluster '0') using the K-means algorithm. The markers of other colors represent neighborhoods that were sorted to different clusters, which tells us that they have venues that are unique.

After I was done with clustering, I sorted the dataframe along with the cluster labels according to the Average Rent, which I then displayed on a bar plot. I found unique neighborhoods that have apartments at a moderate rent amount.

The bar plot with the clustering results is displayed below. The legend describes the different clusters that the neighborhoods are sorted into.



Results and Discussion

As we can see, adding more clusters does not separate the neighborhoods in clusters of multiple sizes, but it singles out unique neighborhoods that have less venues in common with others. Therefore, the neighborhoods that do not fall into the first cluster with label '0' are unique from the others and have venues that may not be commonly found in other neighborhoods.

From the results, we can see that neighborhoods Interbay, Ravenna and South Park have unique venues. Furthermore, we see that the South Park neighborhood has a much higher average rent than other neighborhoods, so its uniqueness has a price.

Interbay and Ravenna both have moderate to low average rent for Seattle. After a brief Google search, I found out that Interbay, as the name says, has a Fisherman's terminal and an extensive railroad network to transport shipped goods to other cities and states. It seems more like an industrial area with some good sites, such as a golf course and bike shops.

However, Interbay and South Park popped out as a unique neighborhood after using just three clusters to sort using K-means. This means that they are more unique than Ravenna in terms of the venues they have.

If it were my decision, I would choose to move to Ravenna, a unique neighborhood in Seattle with affordable rent and cultural diversity.

Conclusion

I used a K-means clustering algorithm to single out unique neighborhoods from a list of 12, and I found 3 such neighborhoods by making 4 clusters. Based on rent, I decided that the two neighborhoods I would consider moving into were Interbay and Ravenna. The results are displayed above, so others who would like to use my data to decide can do so.

However, there are a lot more factors that might go into realistically deciding a neighborhood to stay in. For instance, proximity to work, availability of apartments to rent, and access to highways can also be considered in the decision-making process. It all comes down to the individual use case for the information, and results can therefore not be generalized.

The K-means clustering algorithm is helpful, though, and more data might yield a more general result. To make the results more historically accurate, the data used will need to be updated, too.

Another factor that was not considered was the radius of the search query for the Foursquare API. A radius of 500 meters was used because the exact dimensions of each neighborhood were not known, and the search results were assumed to be general for all venues in the neighborhood. This is very likely the biggest source of error in my analysis, and an inclusive search query will help a lot to accurately judge the uniqueness of venues.

How are these results helpful?

Think about it this way. To visit venues that are common to many neighborhoods in Seattle, you can visit one of the many neighborhoods in Seattle. But wouldn't it be nice to stay in a

neighborhood with unique places to walk to? This might also result in cultural diversity of the area, depending on the venues.

Ravenna has many Southern / Soul food and Mediterranean restaurants, and it has its fair share of coffee shops too. Cuisine from different places on Earth can signify the diversity of taste, or simply the ethnicity of the population living there. Either way, this tells us more about these places in a faraway city, and the venues that make it special.