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 be considering 12 neighborhoods, the average rent at these places, and the uniqueness of venues in a 500-meter radius of each neighborhood’s location.

Neighborhood name and the Average Rent in USD will be 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:

![Text, letter

Description automatically generated]()

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

Table

Description automatically generated

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:

Graphical user interface, text

Description automatically generated with medium confidence

# 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 will start with sorting the neighborhoods into 3 clusters first, and then I will move on to sorting neighborhoods into 4 clusters to see if any bigger clusters are formed. Sorting these neighborhoods into more clusters overfits the data, as we will see neighborhoods that are similar to be split into different clusters, which we do not want.

After I am done with clustering, I will sort the dataframe along with the cluster labels according to the Average Rent, which I will then display on a bar plot. Hopefully, I will find unique neighborhoods that have apartments at a moderate rent amount.

# Results

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.

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 make a decision 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.

## 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.

* Introduction where you discuss the business problem and who would be interested in this project.
* Data where you describe the data that will be used to solve the problem and the source of the data.
* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
* Results section where you discuss the results.
* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
* Conclusion section where you conclude the report.