

Optimal coffee shop location in Seattle, WA

Introduction/Business Problem

My client is looking to expand his business to the North, and he requested to explore the city's neighborhoods and determine the best location for opening a coffee shop.

Data

The problem requires:

- The Seattle neighborhood distribution data – will be collected from the Wikipedia page https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Seattle
- Location data (i.e. latitude and longitude) – will be collected from geopy Python library
- Venues data for each neighborhood – will be collected from Foursquare

Methodology

In the 1st step, we retrieve the Seattle neighborhoods' names with their districts. Since our source is Wikipedia, we clean up the data from square bracket references, alternative names, additional comments, etc. We also rename the columns for easier access. To acquire venues data, we need neighborhoods' locations, i.e. latitude and longitude. We collect it from the geopy's Nominatim. Finally, we visualize the neighborhoods on the map to see their distribution across the city.

In the 2nd step we collect the venues data from Foursquare by neighborhood latitude and longitude within the nearest 500m. We send a GET request, which returns us a .json file, and we retrieve only the most important information to us, i.e. the venue name, location and category. Then we are going to determine the density of each venue category by neighborhood via one-hot encoding for venue's category and calculating mean frequencies for each neighborhood. Afterwards, we can name the top most common venues for all neighborhoods.

In the 3rd step we are going to group the neighborhoods by how common their top venues are, and use machine learning to help us. Since we don't know any prerequisites for neighborhood groups, we are going to use an unsupervised machine learning algorithm – k-means clustering. Firstly, we will determine the number of optimal clusters via an elbow method. Then we are going to run k-means on our venues dataset using the optimal number of clusters we established. Visualizing the clusters will help us assess their distribution across the city. Finally, we are going to examine the clusters and focus on areas that don't have top coffee shops, but might be of potential interest.

Results & Discussion

As a result of our research, we've divided the neighborhoods into 5 clusters. Two of them are of particular interest to us.

One of the clusters consists mainly of park areas with some recreation places to follow. When hanging out in parks people tend to stop by nearby places for some snacks. There already are some food courts and fast food restaurants around, but a small cozy coffee shop might be a breath of fresh air.

Another cluster of interest is the largest one, which covers most active Seattle neighborhoods. There are a lot of places to eat out, including coffee shops being most popular venues around. Since we are looking for a

place which is not crowded with coffee shops, we've removed the aforementioned neighborhoods, and filtered those which have other places to eat among top venues around (e.g. restaurants or bars). These neighborhoods could also potentially be a good place for a coffee shop.

Conclusion

The purpose of this research was to determine the optimal neighborhoods in Seattle to open a coffee shop. By clustering the neighborhoods and exploring their most common venues around, we've discovered 40 neighborhoods, where opening a coffee shop makes most sense. The results of the research are a mere recommendation, and shouldn't be used as an only factor in choosing a location, but rather as a mean of narrowing the city area for further additional exploration regarding rent prices, state of the area, etc.