# 1. Introduction

### Background

I live in Milwaukee and for business I have traveled to the west coast quite a bit, specifically to Portland.

I remember running across this article:  <https://www.salon.com/2014/08/20/milwaukee_is_the_new_portland_partner> and wondering - is this really true? They do feel similar, but what does the data say? Do they have similarities in the make-up of their neighborhoods?

### 1.2 Problem

This project will explore similarities and differences of the distinct make-up of both cities at the neighborhood and city-wide level by describing and clustering neighborhoods by venue types using a K-Means clustering algorithm and exploring the frequencies of occurrence of discrete venue types. We will compare and contrast the results of our findings.

### 1.3 Interest

### I believe this results will be of interest to anyone affiliated with these two cities – what would make someone think these two cities are similar and what are the strong differences that account for the general identity of a city? Anyone thinking of moving or investing in these cities will also find the results interesting beyond normal census data.

2. Data Acquisition and Cleaning

2.1 Data Sources

We will be exploring the neighborhoods using a combination of FourSquare's APIs for venues and locations with a combination of open source datasets to define our neighborhoods.

* <https://en.wikipedia.org/wiki/List_of_neighborhoods_of_Milwaukee> - we will use this list as the index of neighborhoods in Milwaukee
* <https://portlandneighborhood.com/portlandzipcodes> - we will use this data to identity and come up with centroids for our Portland neighborhoods
* <https://developer.mapquest.com/documentation/open/geocoding-api/address/get/> -open source geocoder

2.2 Data Cleaning

# Portland Statistics/Clustering

### Tale of the tape – US Census Statistics

Population: 653,115

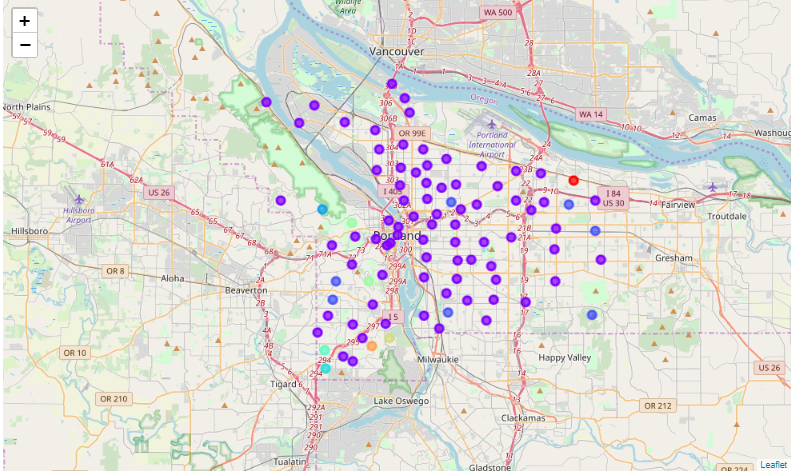
Population / sq mile: 4375

Land area sq mile: 133

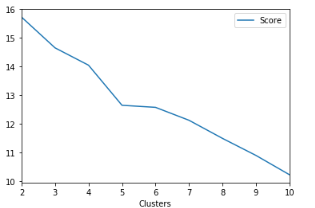
Number of businesses: 75583

Median household income: $61532

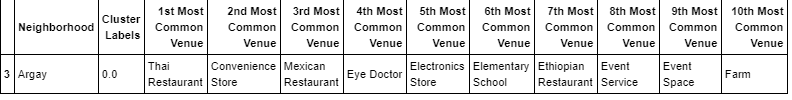
## Portland Neighborhood Clustering – Methodology and Results

To get a baseline of how Portland looks in terms of the venue makeup of its neighborhoods, we will run a K-Means clustering algorithm on a dataset that describes the neighborhoods in the way of venue frequency

We end up with settling on 6 clusters after using the elbow method to determine the ideal number

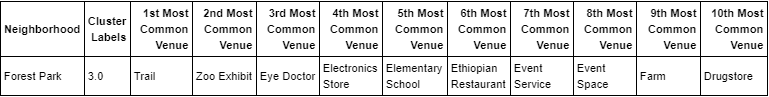


The six clusters created seem to indicate a fairly homogenous dataset.

Cluster 1 simply includes Argay:

Cluster 2 contains 77(!) neighborhoods with nearly no cohesive properties other than coffee shops/bars.

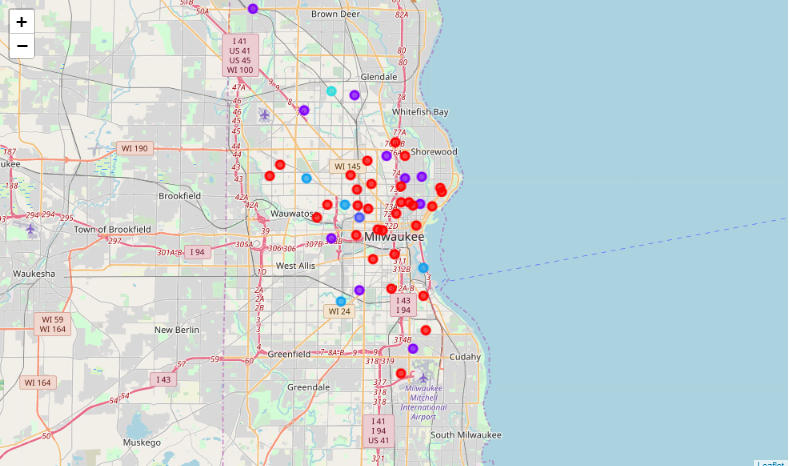
Cluster 3 is very cohesive and contains neighborhoods with heavy Park or Zoo presence. 

Clusters 4,5,6 all contain only a single neighborhood each. 

Portland seems to be very homogenous in terms of overall venue diversity in its’ neighborhoods with some neighborhoods being built specifically to house parklands.

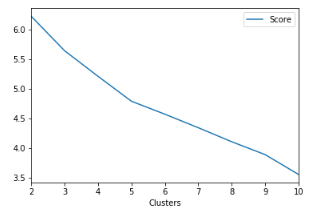
Milwaukee Neighborhood Clustering - Methodology and Results

To get a baseline of how Milwaukee looks in terms of the venue makeup of its neighborhoods, we will run a K-Means clustering algorithm on a dataset that describes the neighborhoods in the way of venue frequency

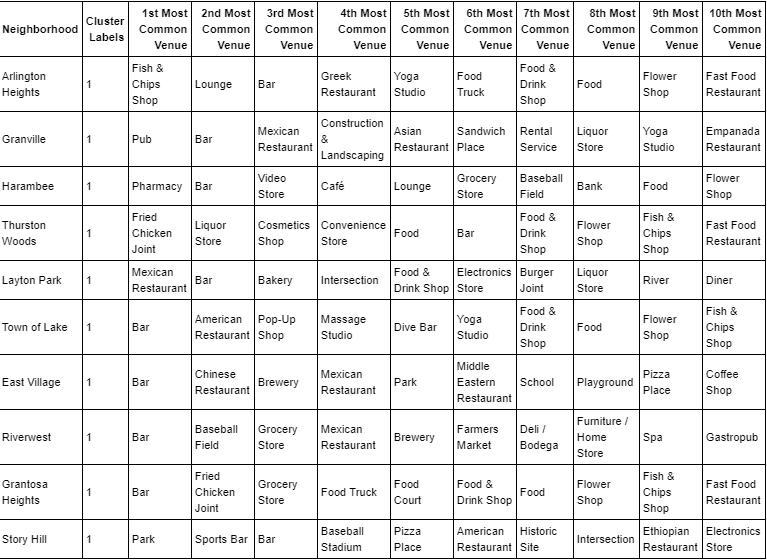


Milwaukee seems to have more distinction in it’s neighborhoods than Portland.

Milwaukee’s data seem best suited for 5 clusters using the elbow method.



Milwaukee had a different feel to it’s clustering than Portland. Cluster 1, while large (30), seemed to be the least harmonious cluster with no real discernable topic. 

Cluster 2 clearly was focused on neighborhoods with high frequencies of bars.

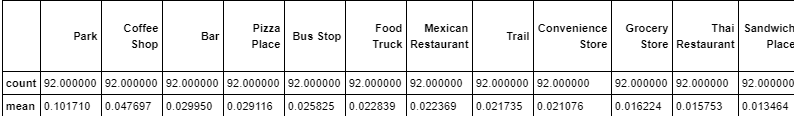
Skipping ahead to Cluster 4, this was clearly focused on parks/playgrounds.

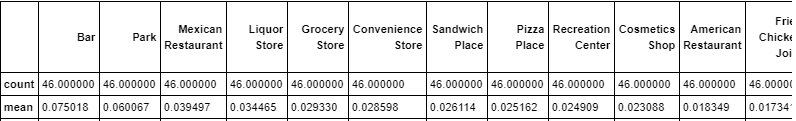
Clusters 3 and 5 stood alone

# Comparison – Discussion and Conclusion

Running a clustering algorithm on a combined data set of PDX and MKE neighborhoods did not show any interesting insight – a single cluster containing 90% of the neighborhoods was created and another cluster of neighborhoods featuring parks was also created. The rest were stand-alone PDX neighborhoods. You could argue that this shows that the cities are similar in that their neighborhoods are very distinctive and have their own flare/flavor – so much so that it is difficult to categorize most of the neighborhoods at all.

I did run a venue frequency analysis on both cities and found that the overall make up of both cities is radically different.

Highest avg frequency PDX venues:

Highest avg frequency MKE venues:

I think a fair conclusion to make here is that the cities are only similar in terms of population and geographic size but in terms of venues and neighborhood make up they are drastically different. There is no evidence to support that these cities share any sort of cultural connection beyond dedication to parks.