

# Capstone Project – The Battle of the Neighborhoods Portland Edition

Dan DeBrock

June 10, 2021

## 1. Introduction

### 1.1 Background

My wife and I are new to the Portland Oregon area and know little about the city. We are big foodies who enjoy cooking and eating a great meal. My wife is originally from Mexico City and has been missing the flavors of home and thinks Portland would be a great place to open a traditional Mexican Restaurant.

### 1.2 Business Problem

I want to analyze the Portland neighborhoods and their venues through clustering and determine the best neighborhood to open a Mexican restaurant.

### 1.3 Interest

This clustering analysis is of interest to my wife and I as we determine where we should spend our capital as well as anyone anyone new to the Portland area.

## 2. Data

### 2.1 Data Sources

The data was collected from 3 different sources.

- Latitude and Longitude data from [public.opendatasoft.com](https://public.opendatasoft.com)
- Postal Code data from [portlandneighborhood.com](https://portlandneighborhood.com)
- Neighborhood venues data from Foursquare

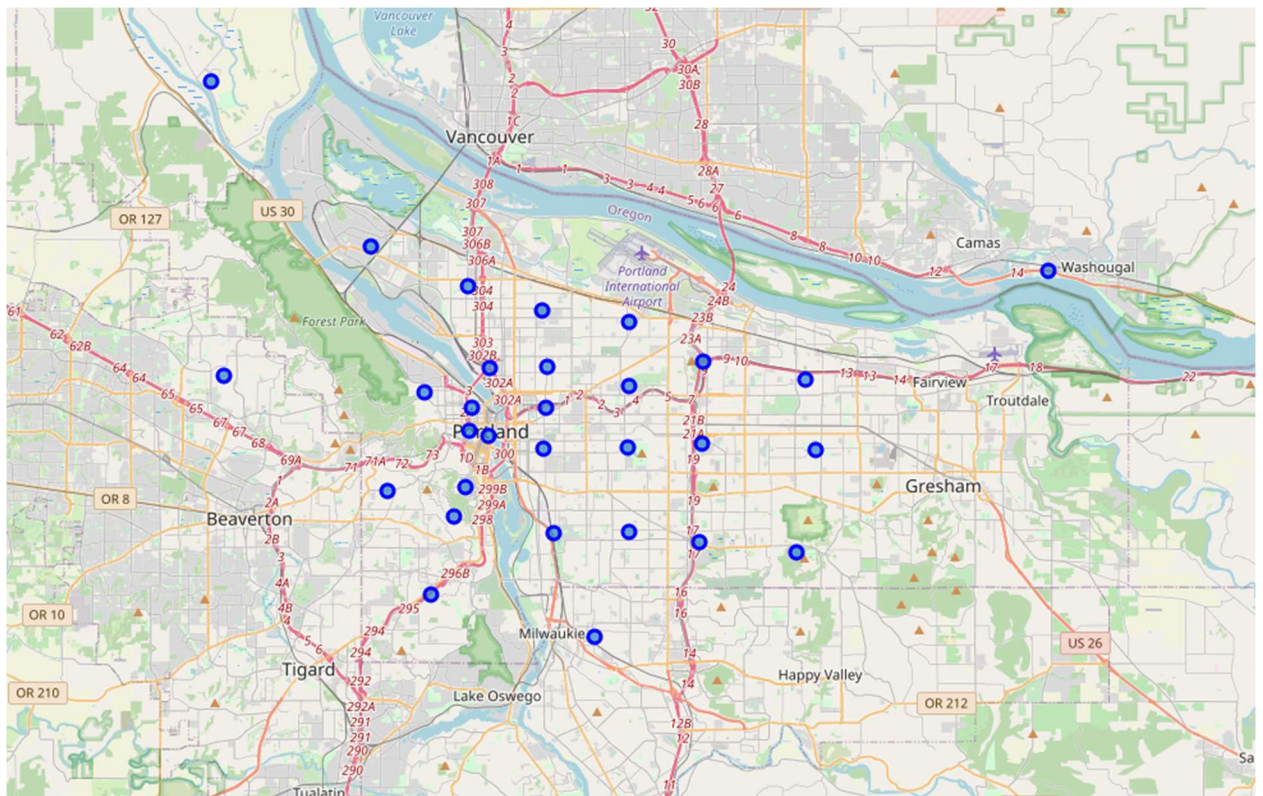
## 2.2 Data Cleansing

The data was downloaded from some sources and scrapped from others. The data was then combined into multiple data frames to analyze the neighborhoods. Some neighborhoods did not return any data from the Foursquare API so those neighborhoods were removed from the analysis.

## 3. Methodology

### 3.1 Exploratory Analysis

Before beginning to explore the venues for each neighborhood I wanted to get a visualization of where the neighborhoods were located around the Portland area.



I then began exploring the venue and venue types in each neighborhood. I used one hot encoding to determine the frequency of each venue type in each neighborhood. I then used that output to determine the Mexican restaurants frequency in each neighborhood. This makes it appear as if

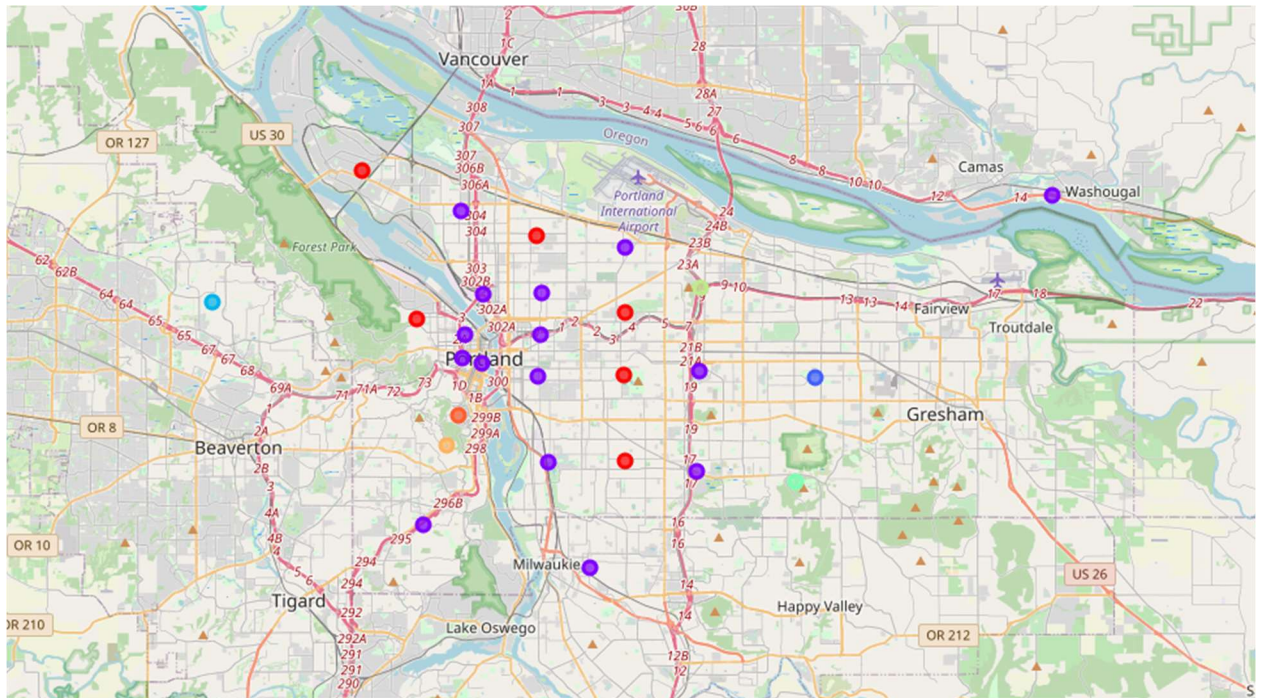
there are many ideal locations to open a Mexican restaurant, but more analysis was needed.

	Neighborhood	Mexican Restaurant
0	Alameda, Beaumont Wilshire, Eliot, Grant Park,...	0.000000
1	Arbor Lodge, Bridgeton, Hayden Island, Humbol...	0.050000
2	Ardenwald	0.111111
3	Arlington Heights, Corbett Terwilliger Lair Hi...	0.000000
4	Arnold Creek, Ash Creek, Collins View, Corbett...	0.000000
5	Boise, Eliot, Overlook	0.153846
6	Brentwood Darlington, Foster Powell, Mt Scott ...	0.000000
7	Brooklyn, Creston Kenilworth, Eastmoreland, Ho...	0.111111
8	Buckman, Hosford Abernethy, Kerns, Laurelhurst...	0.034483
9	Cathedral Park, Portsmouth, St. Johns, Univers...	0.000000
10	Centennial	0.000000
11	Center, Grant Park, Hollywood, Montavilla, Ros...	0.000000
12	Concordia, East Columbia, King, Sabin, Vernon,...	0.000000
13	Cully	0.200000
14	Downtown Portland, Goose Hollow	0.000000
15	Downtown Portland, Goose Hollow, Old Town Chin...	0.010989
16	Forest Park, Northwest Heights	0.000000
17	Hazelwood, Mill Park	0.000000
18	Healy Heights, Homestead	0.000000
19	Hillside, Linnton, Northwest, Portland Northwe...	0.055556
20	Kerns, Laurelhurst, Lloyd District, Sullivans ...	0.000000
21	Lents	0.000000
22	Linnton	0.000000
23	Madison South, Parkrose Heights, Parkrose, Sum...	0.000000
24	Mt Tabor	0.000000
25	Old Town Chinatown, Pearl District	0.000000
26	Pleasant Valley, Powellhurst Gilbert	0.000000
27	Portland Northwest Industrial	0.000000

Using the frequency for each venue type I could determine the top venues in each neighborhood and use this data set for my K-Means Clustering.

### 3.2 K-Means Clustering

I then used the One Hot Encoding table to cluster the neighborhoods based on the top 10 venues for each. I tried using cluster sizes of 5 and 7 to start but they both resulted in one cluster having all neighborhoods minus one per every other cluster. Using 9 clusters brought a divide between the huge cluster from the previous sizes. I then visualized the clusters to see where they were located.



#### 4. Results

The resulting clusters showed the similarities between neighborhoods based on their venues. The first cluster 0 has food trucks as the common venue type. Cluster 1 had the most food venues listed in the top 10 venues from coffee shops, food trucks, and restaurants of all types. The remaining 7 clusters had one neighborhood each. Clusters 2 through 6 venues ranged from farms, offices to parks and mountains but all list a Mexican restaurant as a top 10 venues. The last 2 clusters 7 and 8 only has middle eastern and Mongolian restaurants listed in the top 10.

#### 5. Discussion

Our analysis shows that Portland offers a wide assortment of venues in each neighborhood. Food trucks are huge in Portland and typically there are many in one space, so they are outside of the dense city center. The neighborhoods with the most coffee shops and restaurants are in the heart of the city or right off highways directly from the city. These two clusters are not ideal to open a restaurant let alone a Mexican restaurant. Clusters 2 through 6 all list a Mexican restaurant in their top 10 venues already so I would avoid those areas as well. The two ideal places to open a Mexican restaurant would be clusters 7 and 8 or the Healy Heights, Arlington Heights, Corbett Terwilliger Lair Hill, Hillside, Homestead or Southwest Hills neighborhoods.

## **6. Conclusion**

The purpose of this project was to cluster Portland neighborhoods by venue types in order to aid my wife in narrowing down the search for an optimal location for a new Mexican restaurant. By gathering venue data from Foursquare and identifying the top 10 venues per neighborhood. Clustering was then performed to create subgroups to identify the best neighborhoods to open a Mexican Restaurant.