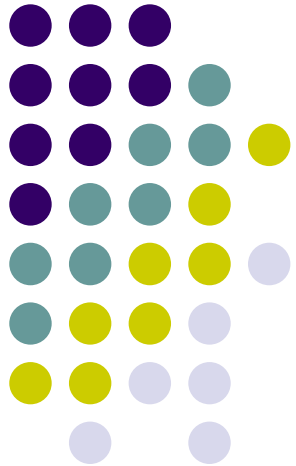
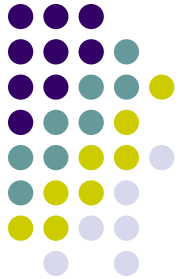
Portland, OR & Sacramento, CA: Compared and Contrasted

By: Mickey McDonell



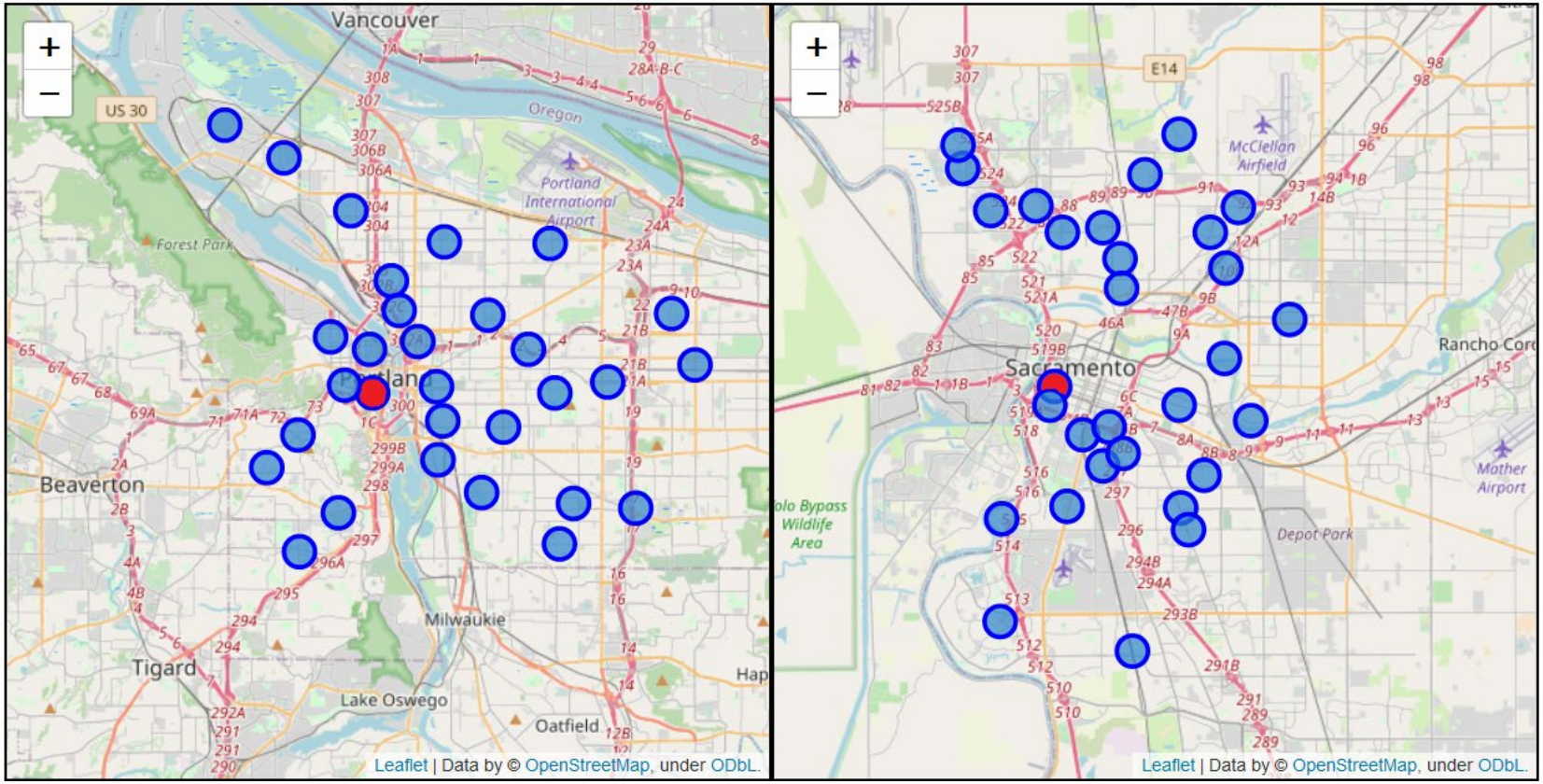
Introduction



- Portland and Sacramento are both West coast cities in the United States with similar populations.
- Portland is the most populated city in the state of Oregon while Sacramento is the capital of California.
- We want to use data analysis and machine learning to better describe these two cities to see how they are similar and different.

Portland

Sacramento



We will be looking at 30 neighborhoods in each city.

*Red dot = Downtown neighborhood

Portland Neighborhoods

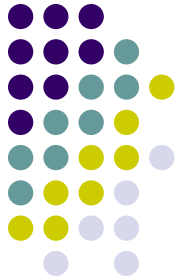
	Name	Population
0	Arbor Lodge	6153
1	Boise	3311
2	Brentwood-Darlington	12994
3	Bridlemile	5481
4	Brooklyn	3485
5	Buckman	8472
6	Cully	13209
7	Downtown	12801
8	Eliot	3611
9	Goose Hollow	6507
10	Grant Park	3937
11	Hazelwood	23462
12	Hillsdale	7540
13	Hosford-Abernethy	7336
14	Lents	20465
15	Lloyd District	1142
16	Montavilla	16287
17	Mount Scott-Arleta	7397
18	Mount Tabor	10162
19	Multnomah	7409
20	Northwest District	13399
21	Parkrose Heights	6363
22	Pearl District	5997
23	Portsmouth	9789
24	Reed	4399
25	Richmond	11607
26	Rose City Park	8982
27	Saint Johns	12207
28	Southwest Hills	8389
29	Vernon	2585

Sacramento Neighborhoods:

	Name	Population
0	Avondale	4485
1	Ben Ali	1359
2	Boulevard Park	2891
3	Colonial Heights	1951
4	Curtis Park	5355
5	Del Paso Heights	4364
6	Downtown	6666
7	East Sacramento	16636
8	Elmhurst	2012
9	Fruitridge Manor	4565
10	Gardenland	3821
11	Gateway West	7864
12	Glenwood Meadows	4894
13	Greenhaven Oaks	9028
14	Hagginwood	9971
15	Land Park	24565
16	Little Pocket	1105
17	Meadowview	26629
18	Midtown	6361
19	Natomas Crossing	3165
20	Northgate	4061
21	Oak Park	14284
22	River Gardens	2118
23	River Park	3558
24	Robla	8182
25	South Natomas	25262
26	Southside Park	2318
27	Sundance Lake	4888
28	Tahoe Park	8398
29	Westlake	3311

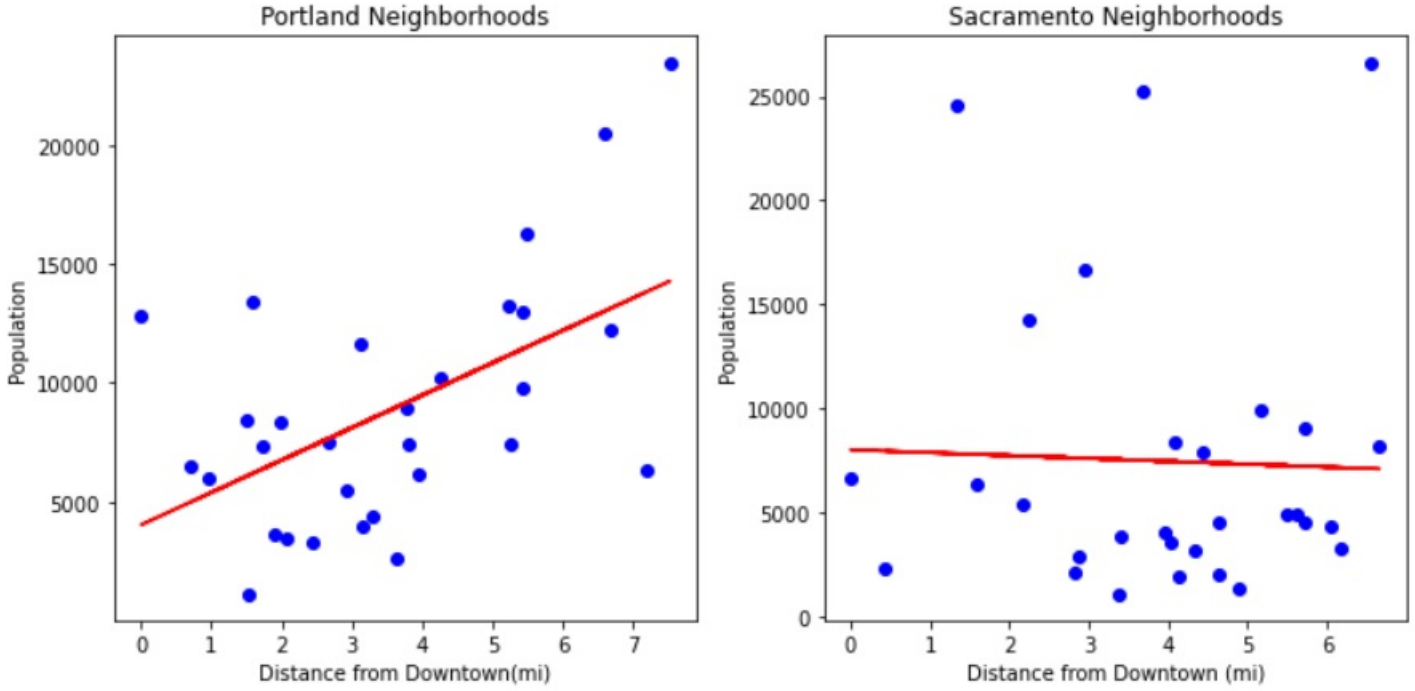
*Please note, population data was gathered from 2010 census data.

Our First Question



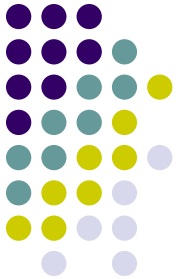
- What is the relationship between a neighborhood's population and distance from downtown for both cities?
- Are the more populated neighborhoods closer to downtown? Or vice versa? Is there even a relationship at all?
- To answer this question, we will be using Linear Regression.

Distance from Downtown and How it Correlates to Population.



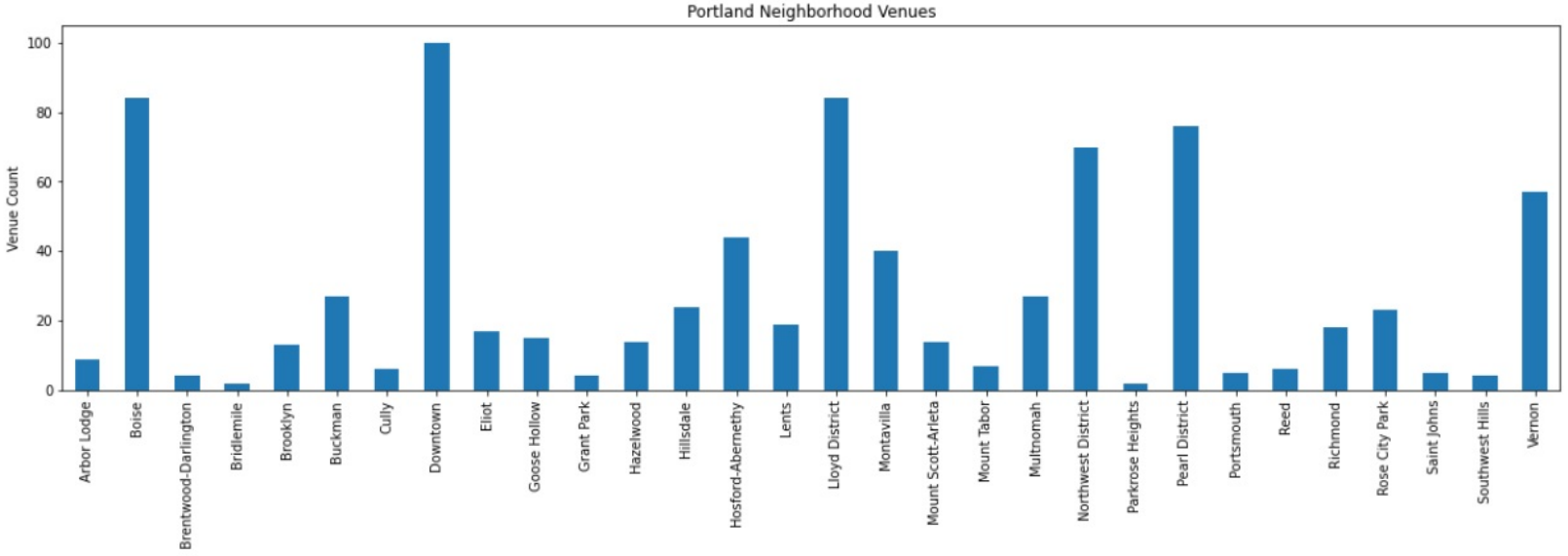
It appears that for Portland, in general, the further the neighborhood is from downtown, the more populated it is. Whereas with Sacramento, there does not seem to be much correlation. Notice the 3 points on the Sacramento graph that hover in the 25,000 range. That would be Land Park, South Natomas and Meadowview. These three neighborhoods are definite outliers.

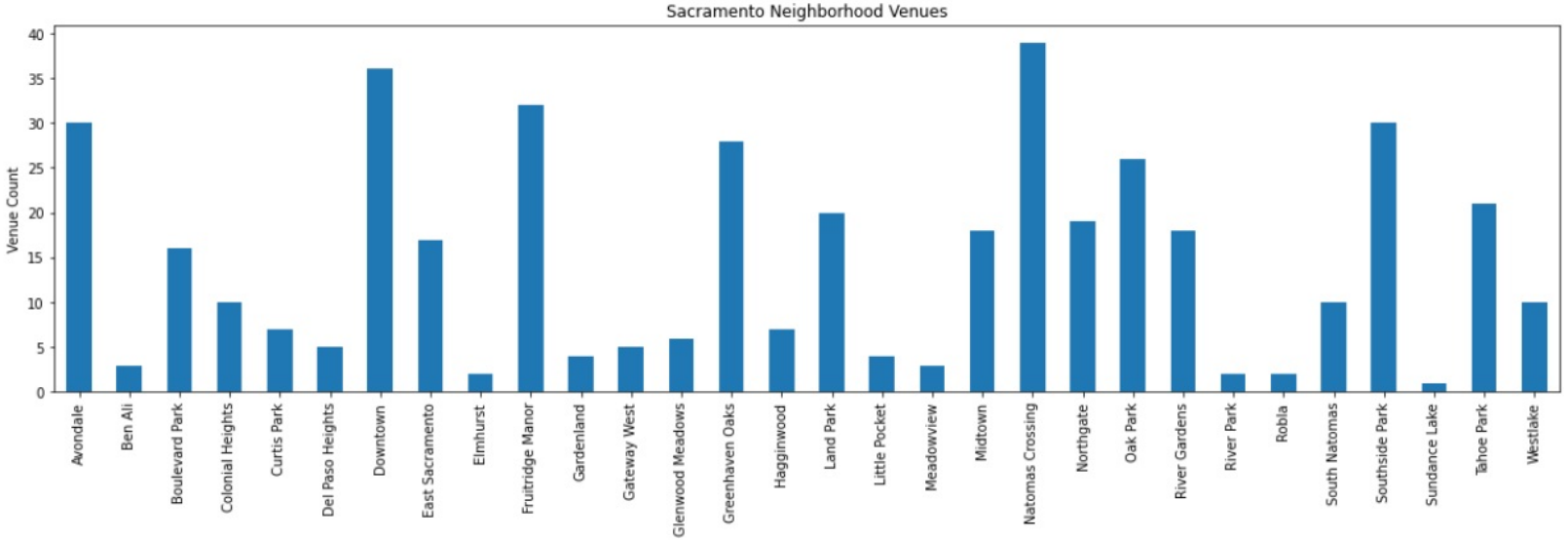
Introducing Foursquare API



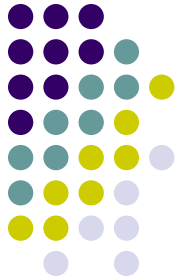
- So far all we have to compare the two cities are population and coordinates, thankfully we can use Foursquare to gather information about local venues, restaurants, shops and attractions within each neighborhood.
- With more data, we will have more comparisons to make.
- Our calls to the Foursquare API will search 500 meters around the neighborhood's coordinates and return up to 100 venues maximum.

Now let's look at the number of venues offered in each city.



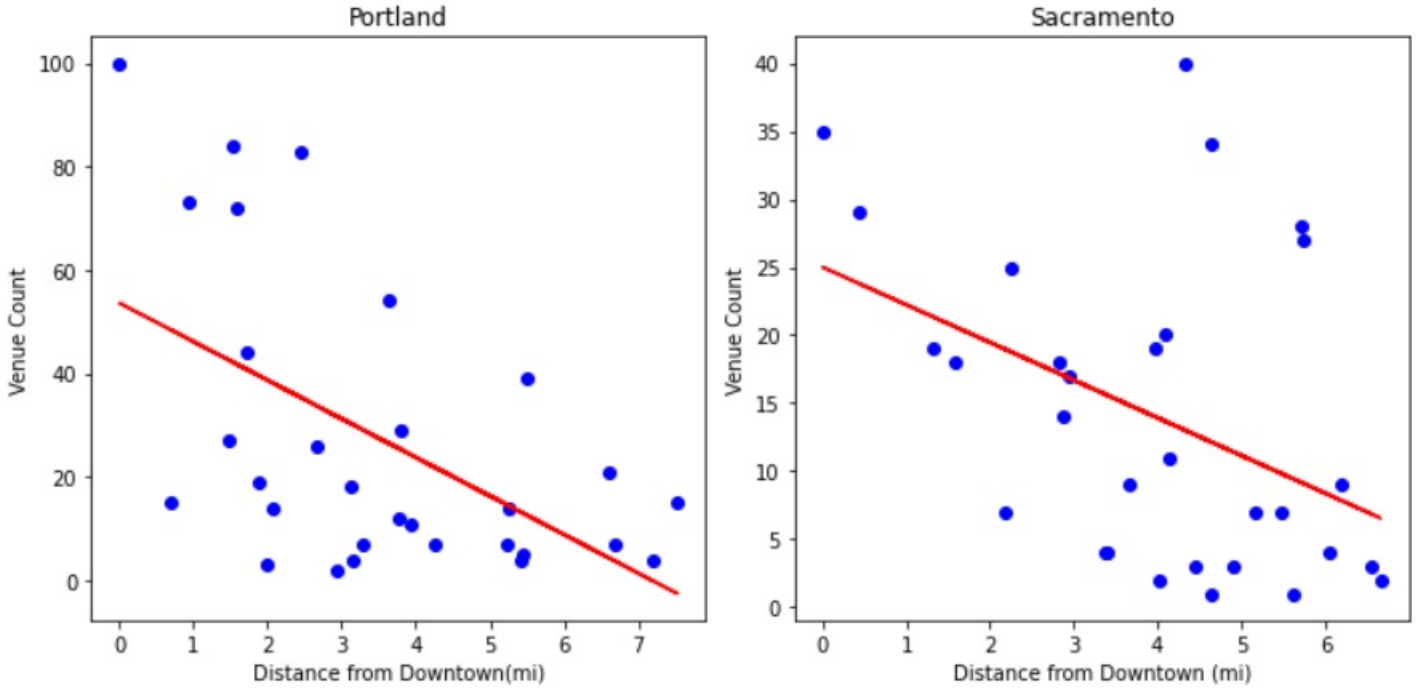


Venue Totals



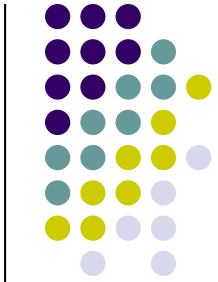
- We gathered 820 venues for Portland from Foursquare.
- And we obtained 431 venues for Sacramento.
- Note that while Foursquare is a great resource, it is unlikely to find every shop, store or attraction in a given range especially for less populated areas where there might not be as much data.
- However it is still interesting to note how many more venues we get in Portland than in Sacramento.

Let's do some more linear regression to see if the venue count and distance from downtown are connected



It seems that, in general, the further we go from downtown, the less venues there are. However, Sacramento's graph seems to have more noise than Portland's (notice the points in the top right corner of Sacramento's graph).

Evaluating our Model

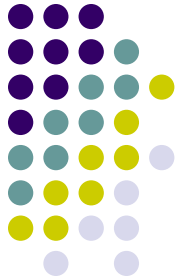


```
Portland's Mean absolute error: 25.91
Portland's Residual sum of squares (MSE): 901.93
Portland's R2-score: 0.31

Sacramento's Mean absolute error: 10.37
Sacramento's Residual sum of squares (MSE): 200.33
Sacramento's R2-score: -0.57
```

- Portland's R Squared score is higher than that of Sacramento's but the MSE is higher as well.
- This is because Portland returned more venues than Sacramento so there was technically more error but because more variance was explained by Portland's model, we get a better R Squared value.

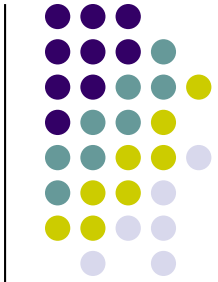
Considering the Most Popular Venue Types



- Now we want to look at the top venues for each neighborhood based on their category and compare them.

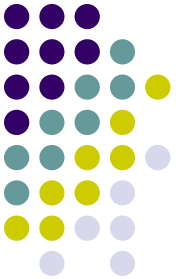
- This will give us a good idea of what kinds of venues are popular in each city.

Preparing a Table for Top Venue Clustering



- First we need the categorical data of each venue type to be rewritten in terms of numbers. Because venue categories cannot be ranked in any meaningful way, we will use one-hot encoding to help our model make sense of the data.

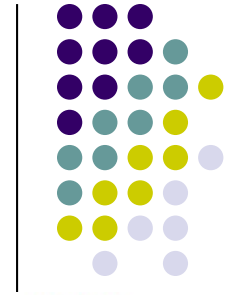
One Hot-Encoding



	Neighborhood	Zoo Exhibit	ATM	Accessories Store	African Restaurant	Airport	American Restaurant	Amphitheater	Art Gallery	Art Museum	...	Tunnel	Vegetarian / Vegan Restaurant	Video Game Store	Video Store
0	Arbor Lodge	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
1	Boise	0.00	0.000000	0.000000	0.000000	0.0	0.011628	0.000	0.000000	0.00	...	0.000000	0.034884	0.011628	0.000000
2	Brentwood-Darlington	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
3	Bridlemile	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
4	Brooklyn	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
5	Buckman	0.00	0.000000	0.000000	0.000000	0.0	0.037037	0.000	0.037037	0.00	...	0.000000	0.037037	0.000000	0.000000
6	Cully	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
7	Downtown	0.00	0.000000	0.000000	0.000000	0.0	0.010000	0.000	0.000000	0.01	...	0.000000	0.020000	0.000000	0.000000
8	Eliot	0.00	0.058824	0.000000	0.058824	0.0	0.058824	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
9	Goose Hollow	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.066667	0.133333	0.000000	0.000000
10	Grant Park	0.00	0.000000	0.000000	0.000000	0.0	0.000000	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
11	Hazelwood	0.00	0.000000	0.000000	0.000000	0.0	0.066667	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.066667
12	Hillsdale	0.00	0.000000	0.000000	0.000000	0.0	0.037037	0.000	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000
13	Hosford-Abernethy	0.00	0.000000	0.000000	0.000000	0.0	0.041667	0.000	0.000000	0.00	...	0.000000	0.020833	0.000000	0.000000

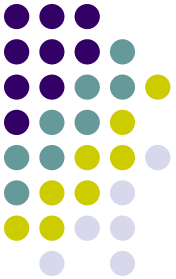
Now the number of venues are represented as a number between 0 and 1 where 0 means there is no venue and the larger the value gets the more types of this specific venue exist in the neighborhood compared to the total number of venues in each neighborhood. You can think of it as a percent.

Let's quickly look at some of Portland's top 10 venues:



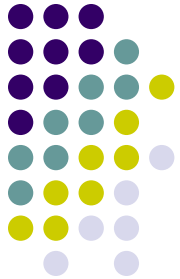
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbor Lodge	Convenience Store	Marijuana Dispensary	Sushi Restaurant	Mexican Restaurant	Coffee Shop	Pizza Place	Bus Stop	Park	Thai Restaurant	NaN
1	Boise	Food Truck	Pizza Place	Coffee Shop	Cocktail Bar	Vegetarian / Vegan Restaurant	Pet Store	Bar	Flower Shop	Café	Gift Shop
2	Brentwood-Darlington	Dog Run	Deli / Bodega	Bus Stop	Park	NaN	NaN	NaN	NaN	NaN	NaN
3	Bridlemile	Tennis Court	Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Brooklyn	Bar	Bus Stop	Farmers Market	Café	Soccer Stadium	Arts & Crafts Store	Grocery Store	Park	Marijuana Dispensary	NaN
5	Buckman	Coffee Shop	Bar	Brewery	Mexican Restaurant	Russian Restaurant	Martial Arts School	Park	Concert Hall	Sandwich Place	Chinese Restaurant
6	Cully	Convenience Store	Marijuana Dispensary	Comedy Club	Farm	Mexican Restaurant	Taco Place	Gas Station	Yoga Studio	NaN	NaN
7	Downtown	Coffee Shop	Hotel	Park	Café	Performing Arts Venue	Pizza Place	Plaza	Food Truck	Sandwich Place	Hotel Bar
8	Eliot	Brewery	Sandwich Place	Gym / Fitness Center	Mexican Restaurant	Bookstore	Park	Tapas Restaurant	Lounge	Gym	Bus Stop
9	Goose Hollow	Pub	Vegetarian / Vegan Restaurant	Pizza Place	Thai Restaurant	Cooking School	Coffee Shop	Soccer Stadium	Sushi Restaurant	Tunnel	Indian Restaurant
10	Grant Park	Park	Pizza Place	Thai Restaurant	Mobile Phone Shop	NaN	NaN	NaN	NaN	NaN	NaN
11	Hazelwood	Pharmacy	Asian Restaurant	Fried Chicken Joint	Fast Food Restaurant	Mexican Restaurant	Shipping Store	Casino	Salon / Barbershop	Pizza Place	Japanese Restaurant
12	Hillsdale	Brewery	Pizza Place	Mexican Restaurant	Pool	Pilates Studio	Pharmacy	Sandwich Place	Farmers Market	Coffee Shop	Thai Restaurant
13	Hosford-Abernethy	Bar	Coffee Shop	Food Truck	Mexican Restaurant	Garden	American Restaurant	Pizza Place	Ice Cream Shop	Breakfast Spot	Grocery Store

And some of Sacramento's top 10 venues:



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Avondale	Sandwich Place	Rental Car Location	Pizza Place	Coffee Shop	Clothing Store	Smoke Shop	Greek Restaurant	Gym	Fast Food Restaurant	Hawaiian Restaurant
1	Ben Ali	Marijuana Dispensary	Train Station	Storage Facility	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Boulevard Park	Discount Store	Bakery	Sporting Goods Shop	Grocery Store	Supermarket	Sushi Restaurant	Coffee Shop	Pharmacy	Convenience Store	Bank
3	Colonial Heights	Grocery Store	Furniture / Home Store	Gym / Fitness Center	New American Restaurant	Mexican Restaurant	Park	Chinese Restaurant	Hotel	Ice Cream Shop	Pizza Place
4	Curtis Park	Park	Furniture / Home Store	Soccer Field	Home Service	Tennis Court	NaN	NaN	NaN	NaN	NaN
5	Del Paso Heights	Bookstore	Shipping Store	Business Service	Liquor Store	Nightlife Spot	NaN	NaN	NaN	NaN	NaN
6	Downtown	Café	Gym	Art Gallery	Asian Restaurant	Farmers Market	Deli / Bodega	Coffee Shop	Park	State / Provincial Park	Gym / Fitness Center
7	East Sacramento	Juice Bar	Pet Store	Dry Cleaner	Breakfast Spot	Eye Doctor	Sushi Restaurant	Pharmacy	English Restaurant	Coffee Shop	Thai Restaurant
8	Elmhurst	Trail	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Fruitridge Manor	Vietnamese Restaurant	Pizza Place	Fast Food Restaurant	Grocery Store	Breakfast Spot	Liquor Store	Gas Station	Donut Shop	New American Restaurant	Sporting Goods Shop
10	Gardenland	Mexican Restaurant	Gas Station	Liquor Store	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	Gateway West	Playground	Home Service	Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	Glenwood Meadows	Pizza Place	Sandwich Place	Clothing Store	Donut Shop	Mexican Restaurant	Shopping Plaza	Grocery Store	NaN	NaN	NaN
13	Greenhaven Oaks	ATM	Pizza Place	Coffee Shop	Department Store	Fast Food Restaurant	Intersection	Lawyer	Nail Salon	Pet Store	Pharmacy

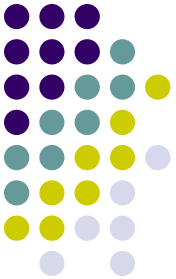
Let's Learn More About Cafés



- It seems like coffee shops, cafés and tea rooms tend to be popular in bustling cities. In the data above, we saw several cafés/coffee shops in the top 10 somewhere.

- Let's see if we can use logistic regression to predict if a neighborhood will have a café.

First we make a table with a column that tells us if the neighborhood has a café or not.



Portland

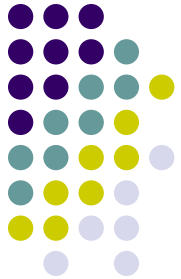
	Name	Latitude	Longitude	Population	Has Cafe
0	Arbor Lodge	45.571794	-122.690152	6153	1
1	Boise	45.550159	-122.671878	3311	1
2	Brentwood-Darlington	45.468707	-122.597633	12994	0
3	Bridlemile	45.492559	-122.726693	5481	0
4	Brooklyn	45.494819	-122.651552	3485	1

Sacramento

	Name	Latitude	Longitude	Population	Has Cafe
0	Avondale	38.597414	-121.397586	4485	1
1	Ben Ali	38.615511	-121.426040	1359	0
2	Boulevard Park	38.532892	-121.495858	2891	1
3	Colonial Heights	38.532407	-121.445509	1951	0
4	Curtis Park	38.547359	-121.480280	5355	0

A ‘1’ means the neighborhood has a café, a ‘0’ means it does not. This is a form of supervised learning because we know ahead of time, if the neighborhood has a café or not.

Now we split the data into a train and test set.



- We will be using neighborhood population and venue count to predict if there is a café or not.

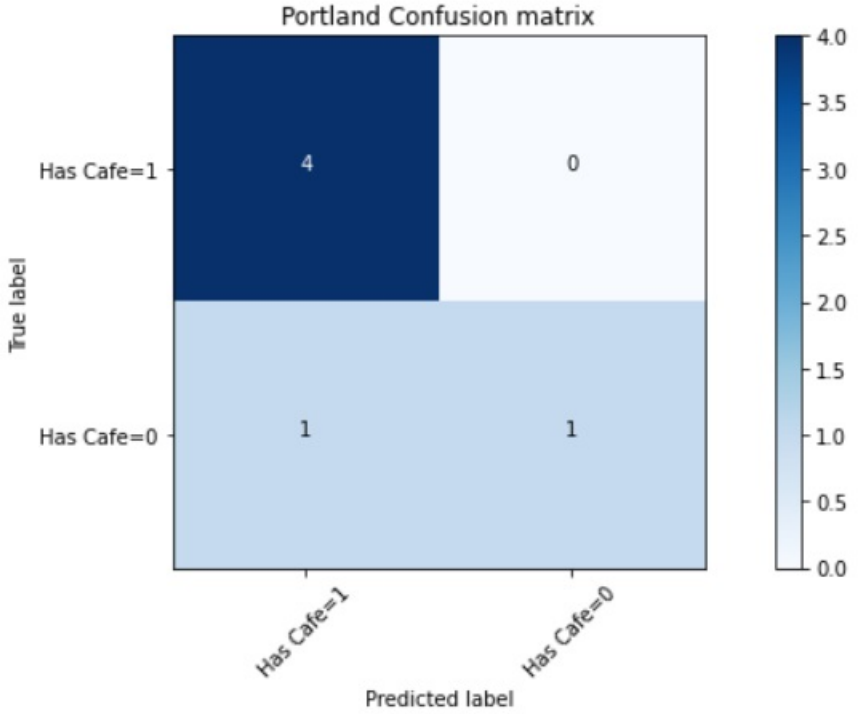
- We have 24 neighborhoods in the training section and 6 in the testing section.

```
X_train, X_test, y_train, y_test = train_test_split( X_log, y_log, test_size=0.2, random_state=4)
print ('Train set:', X_train.shape,  y_train.shape)
print ('Test set:', X_test.shape,  y_test.shape)
```

```
Train set: (24, 2) (24,)
Test set: (6, 2) (6,)
```

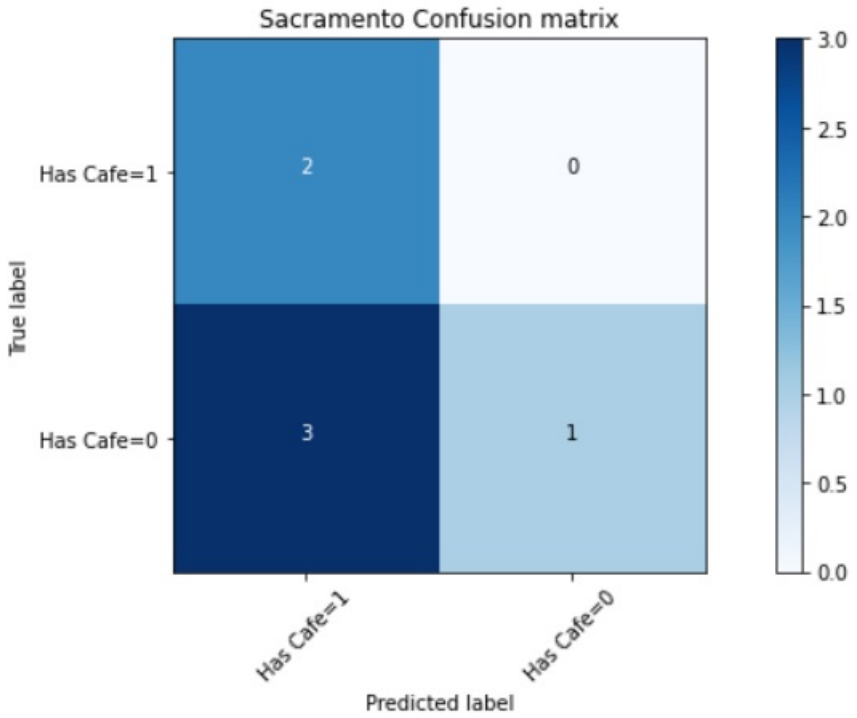
Results

- Our Portland model correctly predicted 4 neighborhoods having a café as true and 1 neighborhood without a café as not having a café. It did misclassify 1 neighborhood that did not have a café as having one. 5/6 correct.

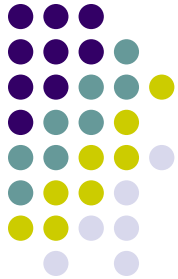


Results (Cont.)

- Our Sacramento model did not perform as well. It mislabeled 3 neighborhoods that did not have a café as having one. It got 2 neighborhoods that did have a café correct and 1 that did not have a café correct as well. 3/6 correct.



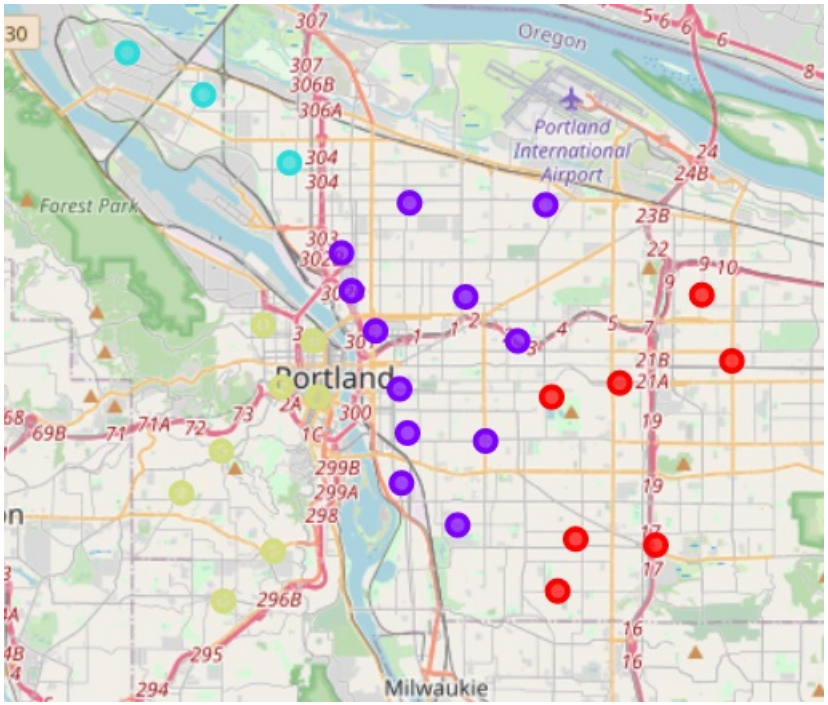
Clustering Neighborhoods



- Finally, let's look at some ways we can meaningfully group the neighborhoods. First, let's see how they can be grouped by distance (longitude, latitude) into 4 clusters.

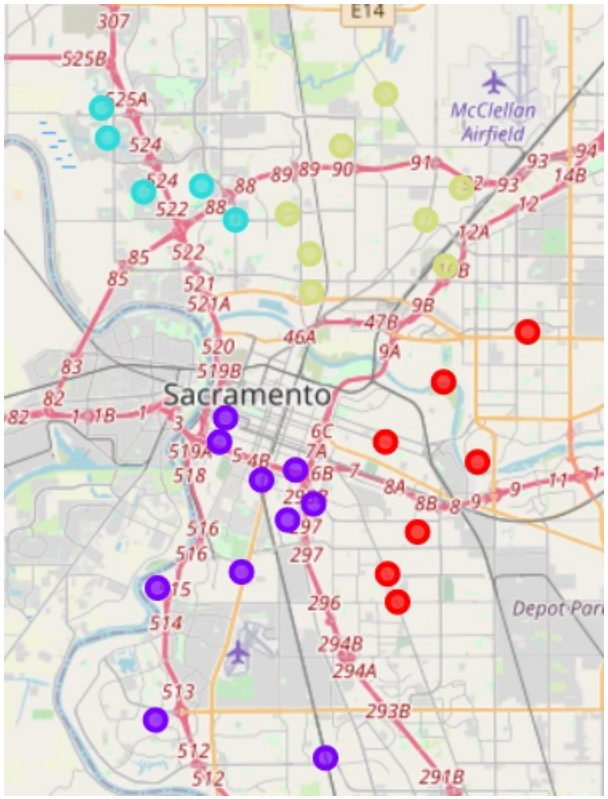
- We will be using K-Means Clustering to achieve this goal. With a K value of 4

Portland



We'll call the cyan dots North Portland, the purple dots Center Portland, the red dots East Portland and the beige dots West Portland.

Sacramento



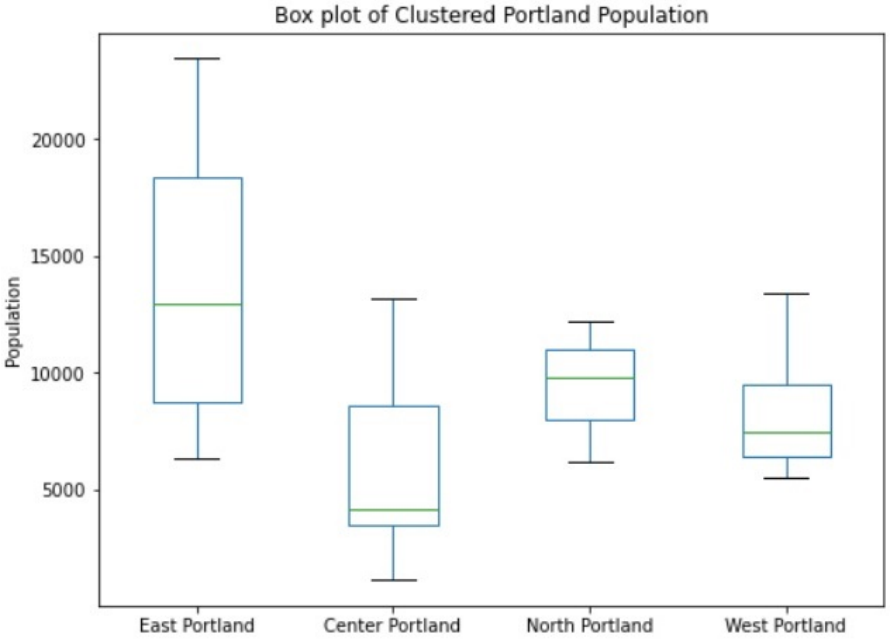
Cyan: Northwest Sacramento

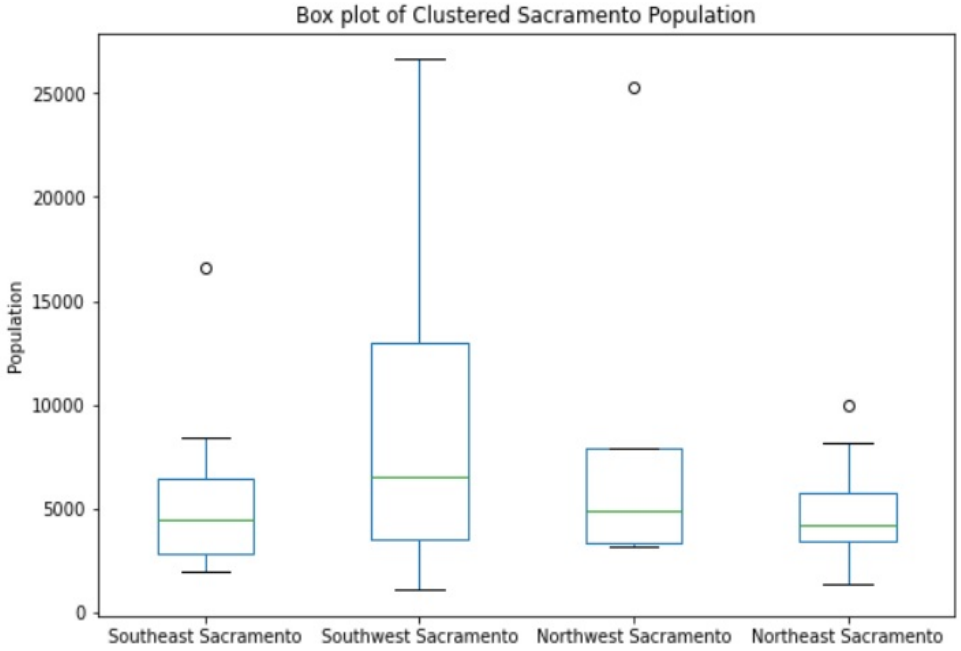
Beige: Northeast Sacramento

Red: Southeast Sacramento

Purple: Southwest Sacramento

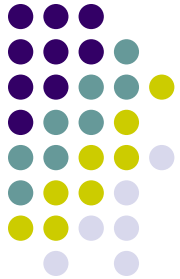
From these groups we can compare the populations in each cluster.





Notice how Sacramento's cluster populations average at around 5,000 each whereas there is more diversity in the Portland populations for the clusters. Sacramento's clusters have more outliers however.

Final K-Means Clustering

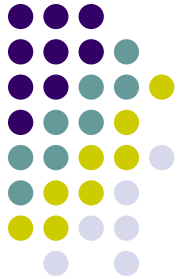


- Let’s do one last K-Means Cluster with the venue information we acquired from Foursquare.

- We will be clustering neighborhoods, not based on location but based on most common venue categories.

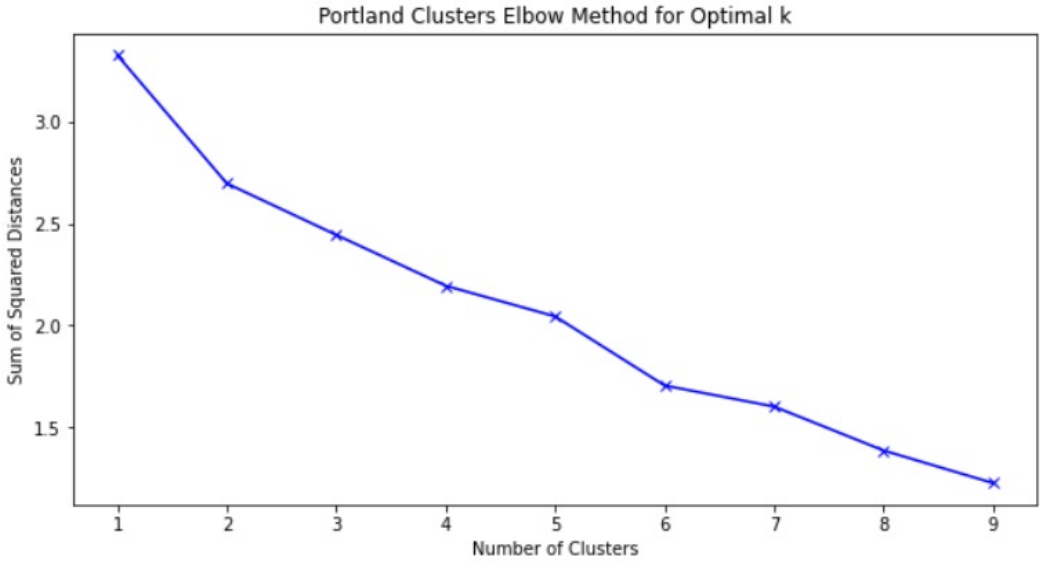
- Let’s use the elbow method to help us decide a reasonable value for K.

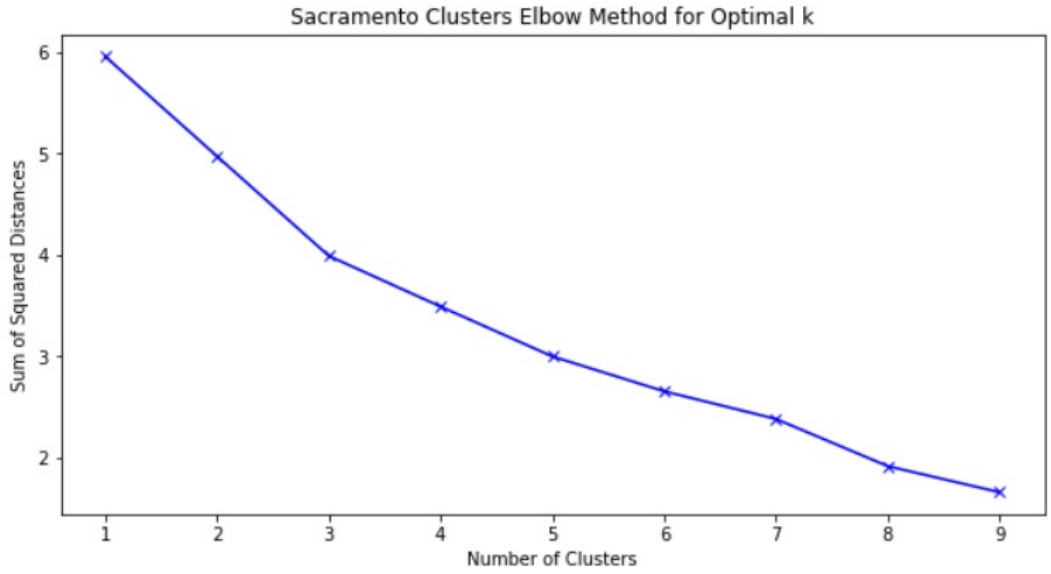
Elbow Method



- With the elbow method, we want to find the cutoff point for our K value. This is determined at the point where adding another cluster group would not reduce the distortion of the model substantially.
- To put it another way, the more groups we make, the more centralized those groups will be, but then we start to take away the insight we get from generalizing unique points into groups in the first place.

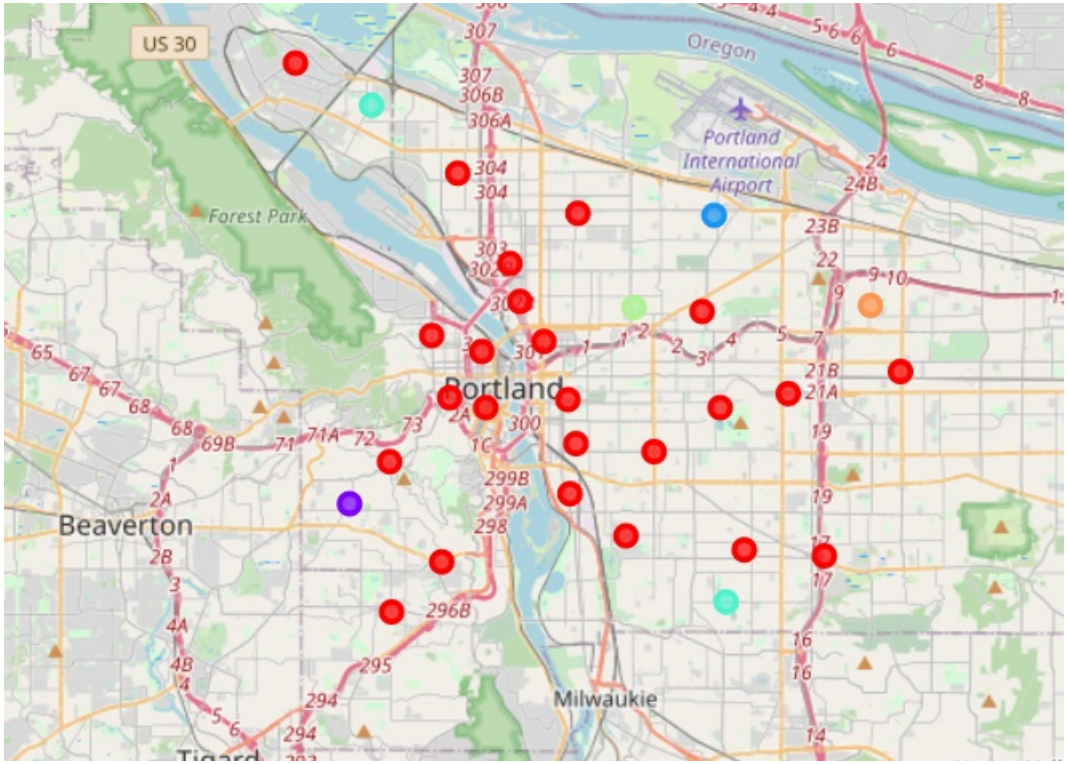
Elbow Method Graphed

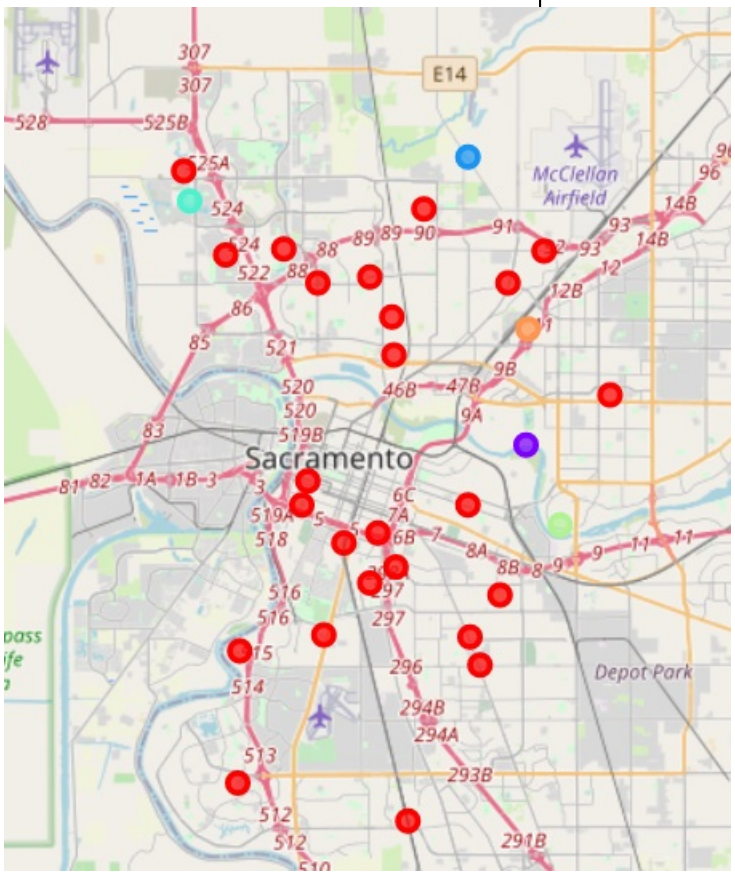




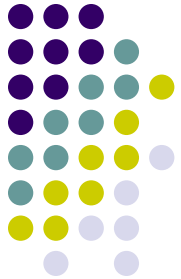
Our graphs look a little wonky, but I have determined that 6, for both cities, is the ideal value for K.

Venue Category K-Means Maps



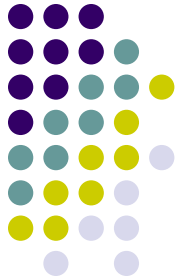


Venue Category K Means Results Discussed



- Both cities yielded similar groupings when using the venue category counts as the criteria for K-Means clustering. Both Portland and Sacramento had one large cluster that consisted of most neighborhoods.
- This large cluster encapsulated all central neighborhoods in their respective cities as well.
- Another take away is that each of the clusters with only one or two neighborhoods had less than 10 venues returned by Foursquare.
- Some neighborhoods with less than 10 venues did end up in the big cluster within each city, but the venues in these neighborhoods contained popular venue categories like parks, restaurants, food trucks, cafés, convenience stores, etc.

Conclusion



- It seems that Portland has more venues than Sacramento as well a positive linear distribution of neighborhood populations as we move further and further from downtown. Portland's neighborhoods are also spread out further apart than Sacramento's which bunch up more toward the center.
- Sacramento's population did not reliably change much the further we went from downtown, and when splitting the neighborhoods into 4 smaller groups, the average population of each group was very homogenized. There was, however, one neighborhood in each group that had a population much higher than the rest. It seems that most of Sacramento's neighborhood population is found around the 5,000 range with a few noticeable outliers in the 25,000 range.
- The total number of venues in Sacramento was much smaller (nearly half that) of Portland which is quite surprising considering that their populations are not that far apart. This could be due to Foursquare's API missing some venues or also the fact that Sacramento is a state capital and perhaps more jobs and building locations are dedicated to government projects rather than commerce.
- This concludes the Portland, Sacramento project, thank you for reading.

Sources

- Portland population data: https://www.portlandoregon.gov/civic/56897
- Sacramento population data: https://www.cityofsacramento.org/-/media/Corporate/Files/GIS/Maps/Neighborhoods_2010Pop_11x17.pdf?la=en
- Foursquare API: https://developer.foursquare.com/
- Google Maps: https://www.google.com/maps