



John Williams

Educated Relocation

Applied Data Science Capstone – IBM - Coursera

Capstone Presentation

Introduction

- Our hypothetical client is planning a move to the Pacific Northwest
- Moving can be hard
- They want to land in a neighborhood they'd enjoy
- They don't know anything about the neighborhoods in the Pacific Northwest



Intro- cont'd

- They do know two neighborhoods they've enjoyed previously
 - Hampden, Baltimore, MD

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
Hampden	Pizza Place	American Restaurant	Food Truck	Italian Restaurant	Bar	Bakery	Gift Shop	Art Gallery	Record Shop	Boutique

- Decatur, GA

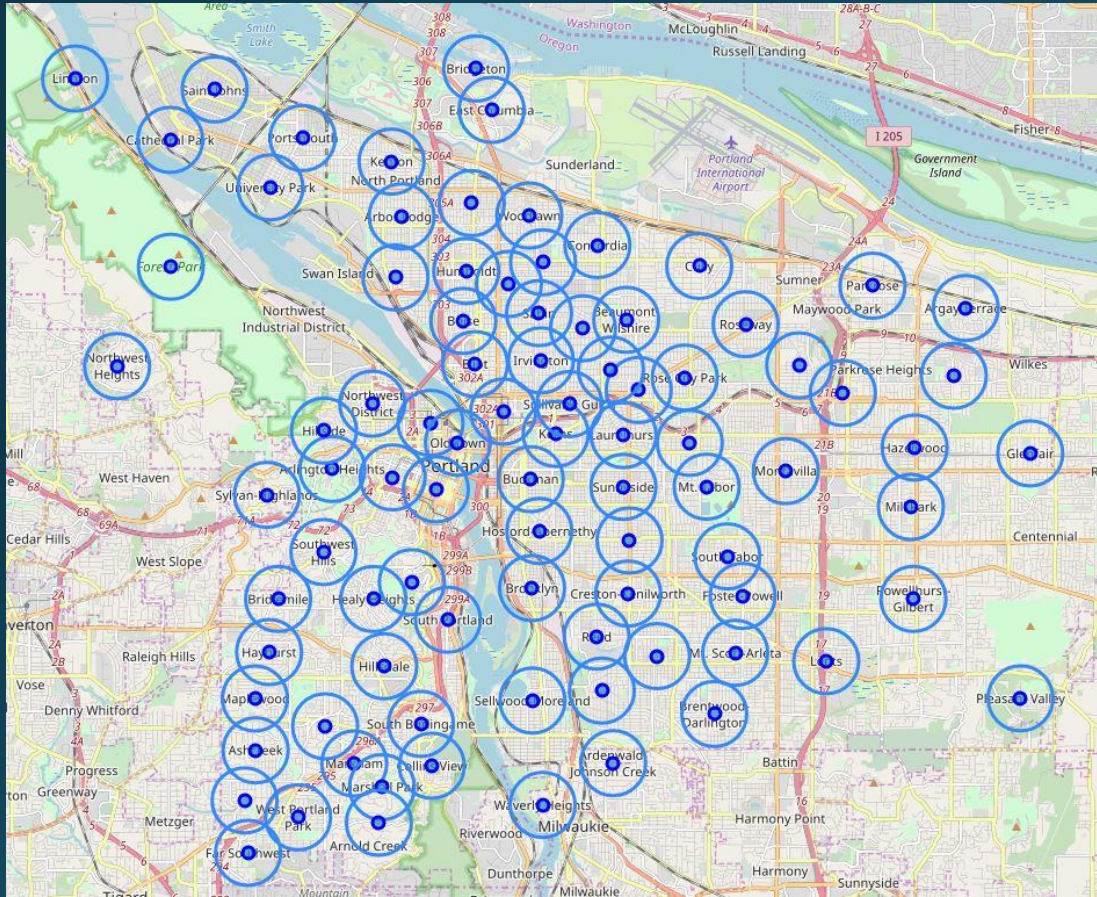
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
Decatur	Pizza Place	Gastropub	Pub	American Restaurant	Coffee Shop	Burger Joint	Arts & Crafts Store	Spa	South American Restaurant	Breakfast Spot

- Based on venue types in the known neighborhoods, we will try to find potential neighborhoods that they'd enjoy

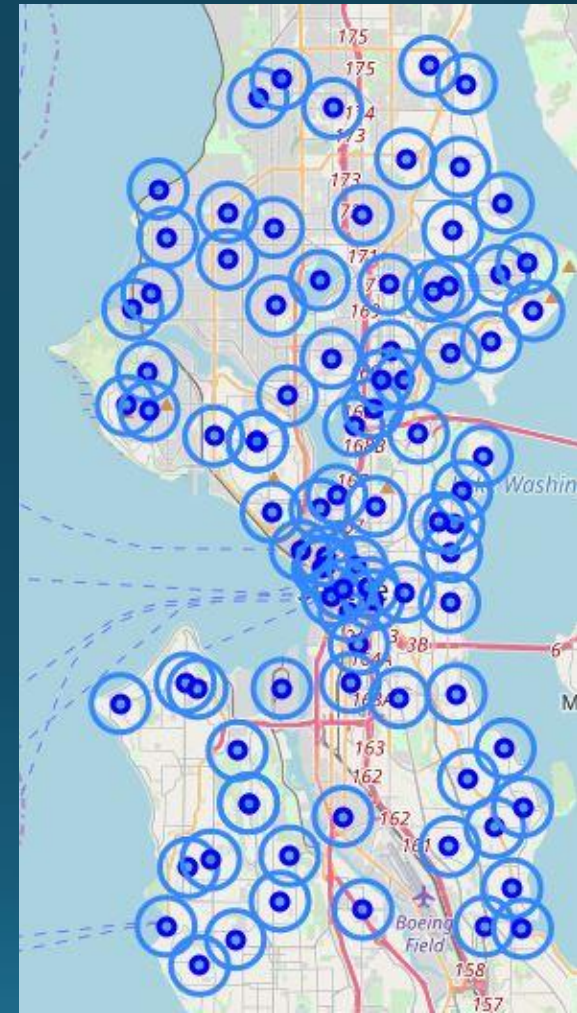
Data

- opendata.arcgis.com : Generate a list of neighborhoods in Portland and Seattle
 - Geolocator to determine neighborhood coordinates
 - Folium to visually inspect for adequate neighborhood coverage
- [Foursquare Places API](#) : generate a list of venues in each neighborhood and further explore select venues to determine rating

Neighborhood Maps



Portland, OR



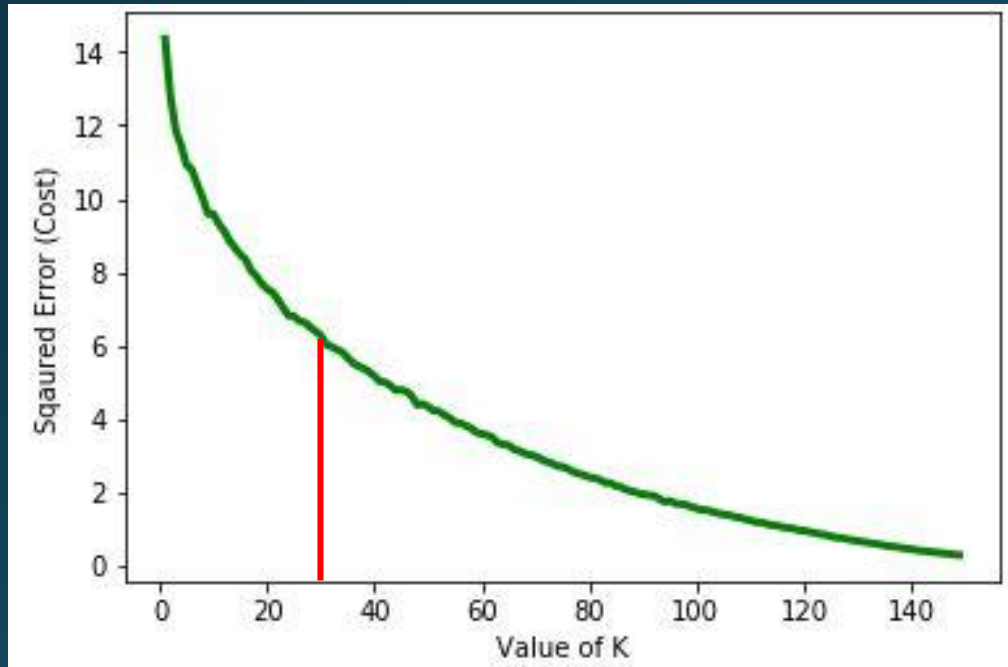
Seattle, WA

Method

- Clustering
 - Use K Means clustering to determine which neighborhoods are most similar to Hampden and Decatur
- Narrowing
 - Use client preferences to pick neighborhoods with a park and a grocery store
- Ranking
 - Rank each neighborhood on grocery store rating

K Means – Elbow Point

Number of clusters versus Squared Error



30 Clusters looked like a reasonable elbow point



Results

Clustering

- Neighborhoods were assigned to 30 clusters

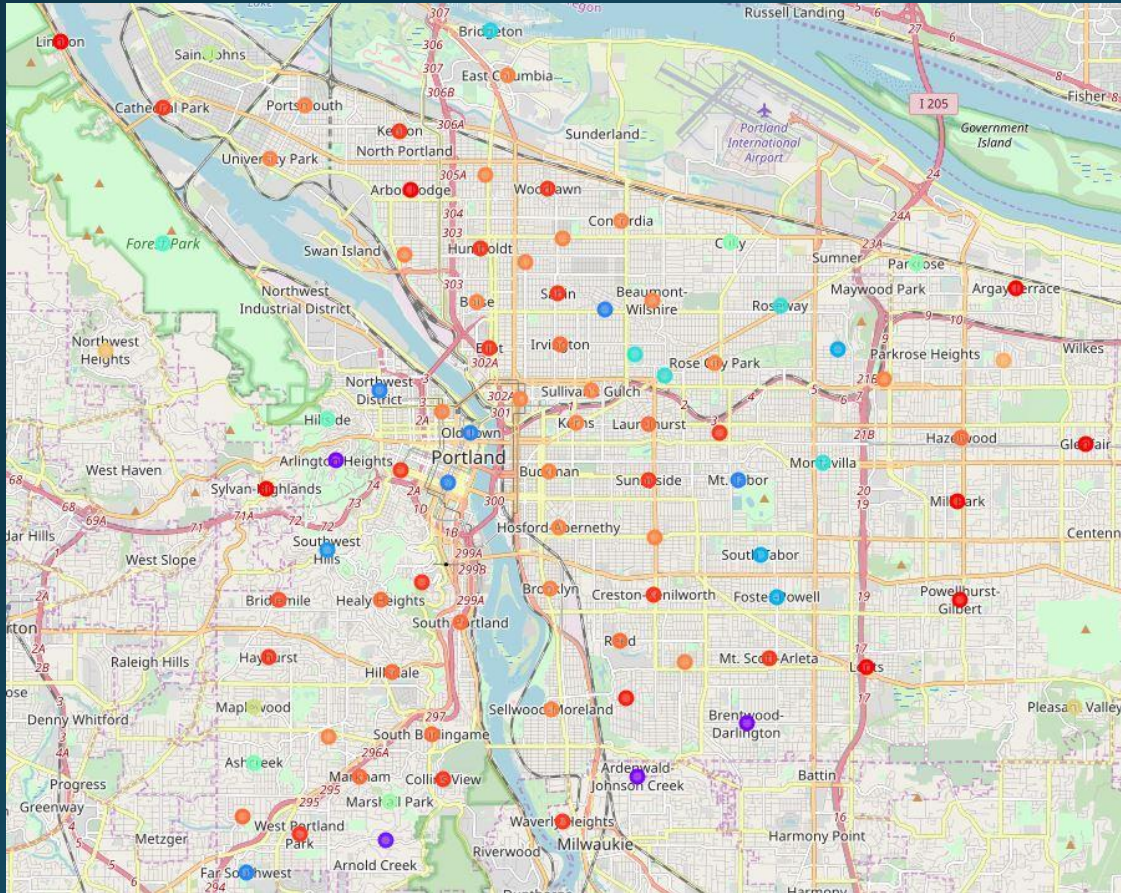
Neighborhoods per Cluster

Cluster Labels	Neighborhood
28	35
25	33
26	26
6	14
8	13
0	11
1	10
11	10
10	4
7	2

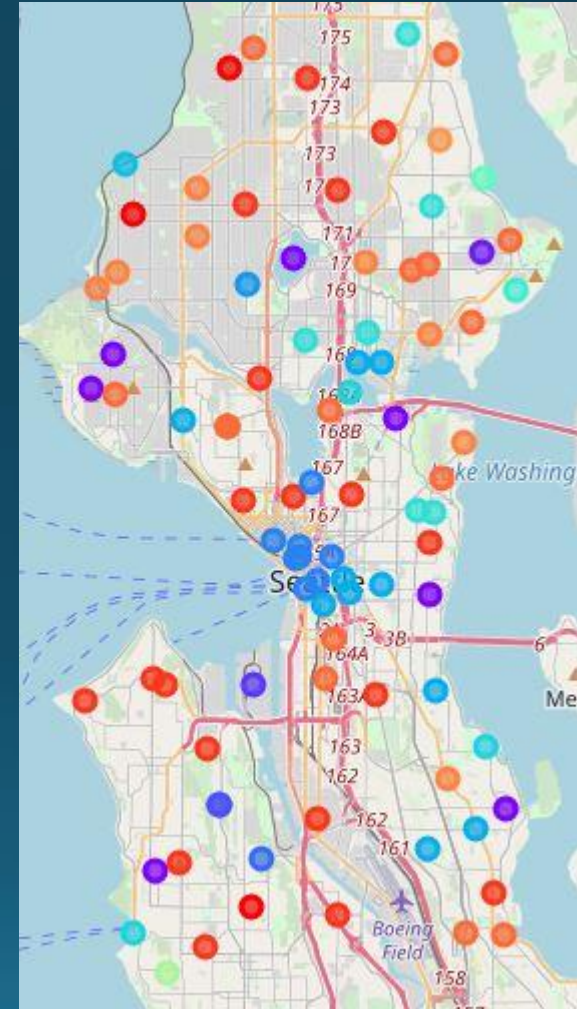
- Hampden and Decatur were in Cluster 25



Mapped Clusters



Portland, OR



Seattle, WA

Cluster Characteristics

- Hampden and Decatur are in Cluster 25

Top 10 Venue Categories in Cluster 25

Venue Category	Neighborhood
Coffee Shop	87
Bar	86
Pizza Place	72
Food Truck	48
Brewery	38
Mexican Restaurant	37
American Restaurant	32
Café	32
Park	32
Thai Restaurant	29

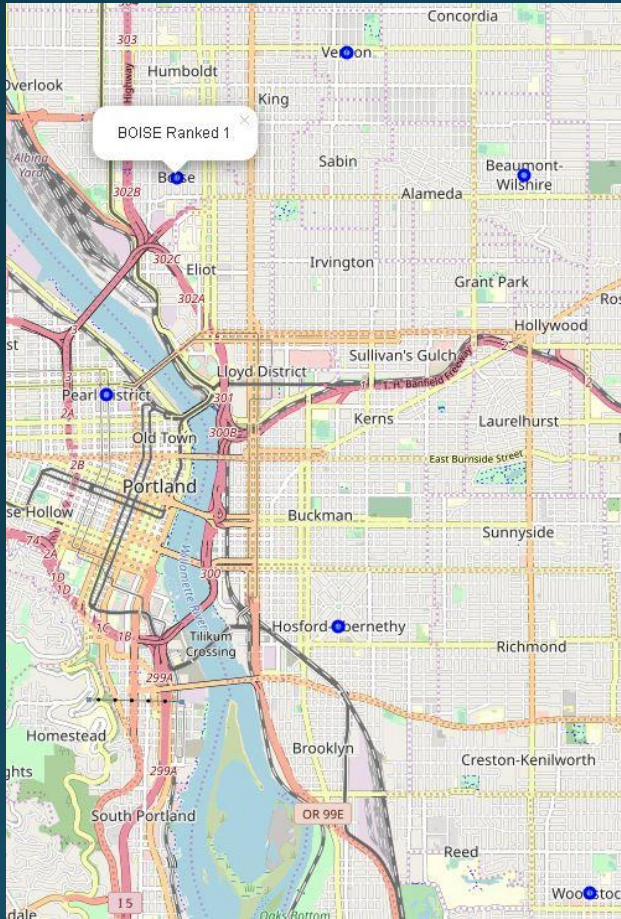
Top 10 Venue Categories in Final 8 Neighborhoods

Venue Category	Neighborhood
Coffee Shop	29
Bar	27
Pizza Place	26
Food Truck	15
Café	14
Brewery	13
Mexican Restaurant	12
Park	12
Breakfast Spot	12
Grocery Store	11

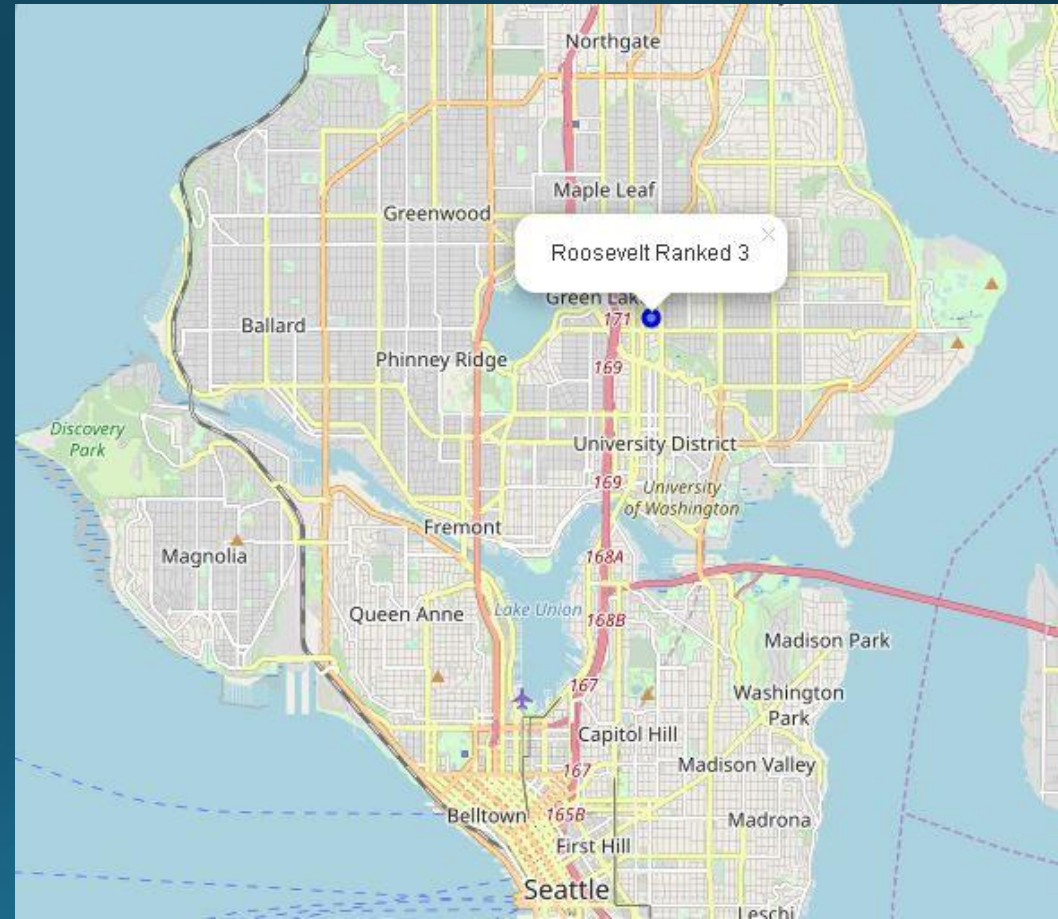
Final Neighborhood Rank

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Rating	Neighborhood Rank
BOISE	45.550159	-122.671878	9.1	1
Hosford-Abernethy	45.506644	-122.649419	8.7	2
Roosevelt	47.677305	-122.313807	8.4	3
VERNON	45.562300	-122.648314	8.2	4
PEARL DISTRICT	45.529044	-122.681598	8.1	5
WOODSTOCK	45.480705	-122.614549	7.8	6
BEAUMONT-WILSHIRE	45.550391	-122.623694	7.1	7
PIEDMONT	45.574552	-122.669405	0.0	8

Final Neighborhood Map



Portland, OR



Seattle, WA

Final Neighborhood Ranking

1. Boise, Portland, OR
2. Hosford-Abernethy, Portland, OR
3. Roosevelt, Seattle, WA
4. Vernon, Portland, OR
5. Pearl District, Portland, OR
6. Woodstock, Portland, OR
7. Beaumont-Wilshire, Portland, OR
8. Piedmont, Portland, OR



Discussion/Limitations

- K Means might not have been the ideal clustering method as no obvious elbow point was present. However, we were able to accomplish the project goal.
- Number of venues per neighborhood was limited to 100.
- Instead of starting with neighborhoods, we could have used sectors to ensure total coverage without significant overlap.
- Limited utility for other clients.

Conclusion

- Using venue category in each neighborhood to cluster similar neighborhoods and narrow the list by ensuring each potential neighborhood has a park and a grocery store, the clients now have a list of 8 neighborhoods ranked by grocery store quality to direct their search.