Capstone Project – Battle of the Neighborhoods

Introduction and Business Problem

- Rapid growth in population of Seattle
  - 2010 – 2018: 22.4% growth in Seattle, 6.0% in U.S.
- Important for business owners to understand types of businesses in area
- Restaurant owners may want to avoid competition from similar restaurants
- May also want to consider median income to determine prices for menu

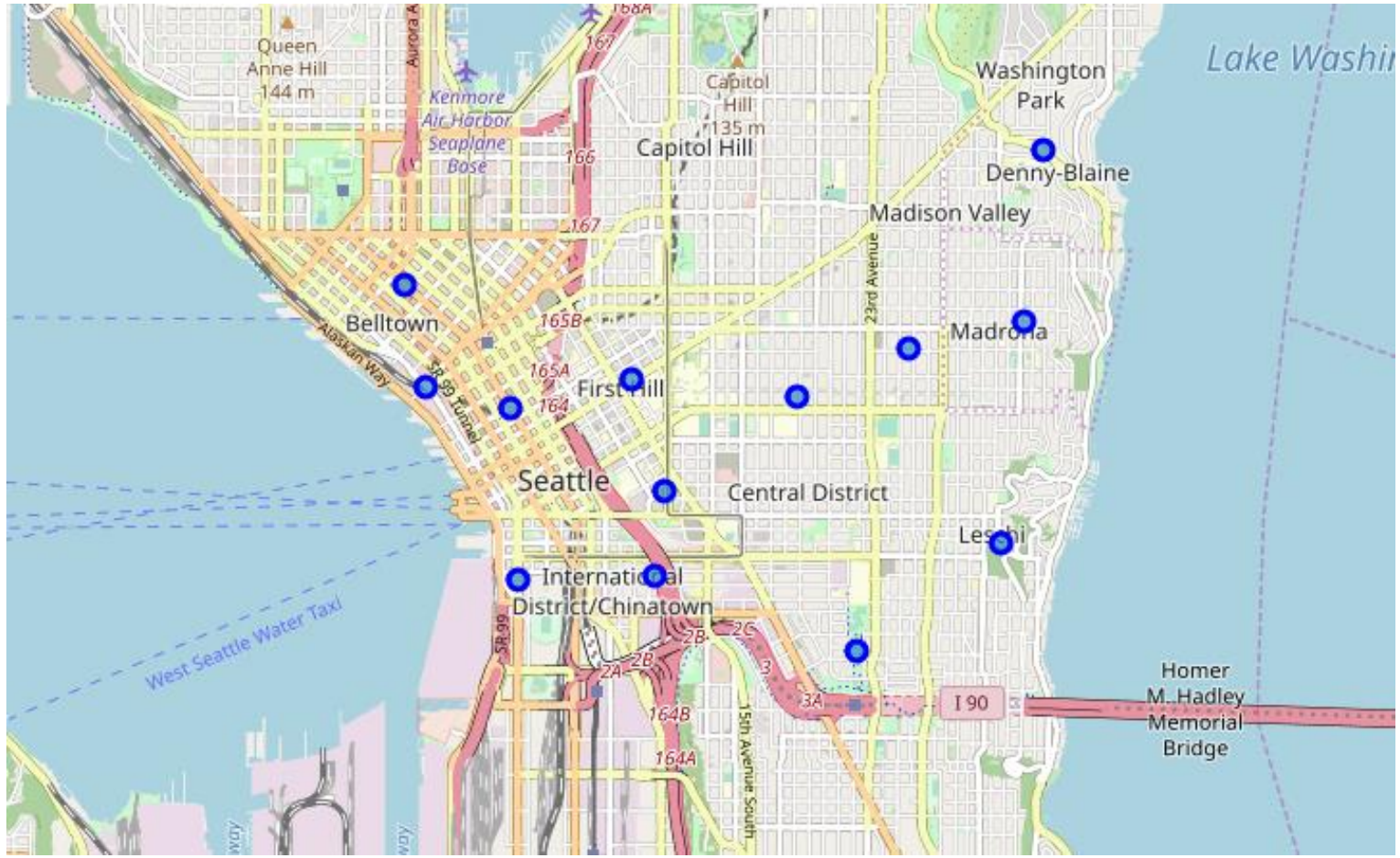
Method

- Focus on Downtown and Central Area districts of Seattle
- Use Foursquare API data to identify most common venues in each neighborhood
  - K-means cluster analysis
- Examine median income of different neighborhoods
  - Generate choropleth map

Data

- GeoJSON file of Seattle
  - Neighborhood
  - Latitude
  - Longitude
  - District
- CSV file of Median Income from various sources
  - Neighborhood
  - MedianIncome

Map of Seattle Neighborhoods



Number of Venues

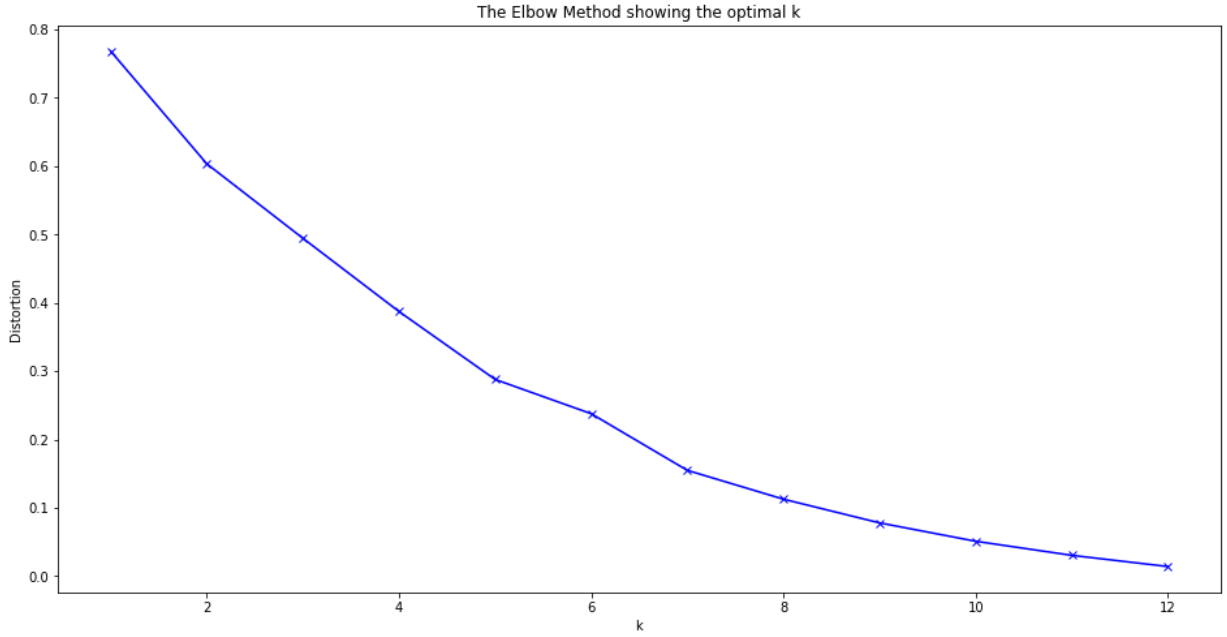
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Atlantic	11	11	11	11	11	11
Belltown	50	50	50	50	50	50
Central Business District	50	50	50	50	50	50
First Hill	50	50	50	50	50	50
Harrison/Denny-Blaine	13	13	13	13	13	13
International District	50	50	50	50	50	50
Leschi	7	7	7	7	7	7
Madrona	18	18	18	18	18	18
Mann	28	28	28	28	28	28
Minor	11	11	11	11	11	11
Pike-Market	50	50	50	50	50	50
Pioneer Square	50	50	50	50	50	50
Yesler Terrace	20	20	20	20	20	20

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	Atlantic	Bus Station	Rental Service	Tunnel	South American Restaurant	Park	Skate Park	Scenic Lookout	Trail	Café	Residential Building (Apartment / Condo)
1	Belltown	Sushi Restaurant	Bakery	Restaurant	Cocktail Bar	Coffee Shop	Italian Restaurant	Bar	Convenience Store	Rock Club	Pet Store
2	Central Business District	Coffee Shop	Hotel	New American Restaurant	American Restaurant	Concert Hall	Clothing Store	Theater	Spa	Music Venue	French Restaurant
3	First Hill	Sandwich Place	Coffee Shop	Lounge	Hotel	Asian Restaurant	Bakery	Bar	Brewery	Italian Restaurant	Nightclub
4	Harrison/Denny-Blaine	Dry Cleaner	Pizza Place	Café	Gas Station	Surf Spot	Sushi Restaurant	Spa	Park	Gym	Bakery
5	International District	Chinese Restaurant	Vietnamese Restaurant	Bakery	Dessert Shop	Hotpot Restaurant	Café	Bubble Tea Shop	Japanese Restaurant	Dive Bar	Dumpling Restaurant
6	Leschi	Park	Playground	American Restaurant	Grocery Store	Pet Store	Pizza Place	Yoga Studio	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant
7	Madrona	Gift Shop	Coffee Shop	Thai Restaurant	Playground	Italian Restaurant	Cupcake Shop	Seafood Restaurant	French Restaurant	Tennis Court	Ethiopian Restaurant
8	Mann	Ethiopian Restaurant	Coffee Shop	Deli / Bodega	Food Truck	Bar	Taco Place	Mini Golf	Donut Shop	Dessert Shop	Fried Chicken Joint
9	Minor	Coffee Shop	Bus Stop	Playground	Food Truck	Soccer Stadium	Theater	Convenience Store	Indie Theater	Art Gallery	Fish Market
10	Pike-Market	Hotel	Cocktail Bar	Coffee Shop	Breakfast Spot	New American Restaurant	Italian Restaurant	French Restaurant	Bakery	Seafood Restaurant	Deli / Bodega
11	Pioneer Square	Coffee Shop	Cocktail Bar	Café	Italian Restaurant	Sporting Goods Shop	Lounge	Mexican Restaurant	Thrift / Vintage Store	Dumpling Restaurant	Pool Hall
12	Yesler Terrace	Vietnamese Restaurant	Noodle House	Korean Restaurant	Thai Restaurant	Poke Place	Sandwich Place	Convenience Store	Food & Drink Shop	Chinese Restaurant	Garden

K-Means Clustering

- Commonly used unsupervised learning method
- Groups observations together
- Elbow method used to determine optimal k-clusters
  - No obvious optimal number of clusters, but slight bend around k=5



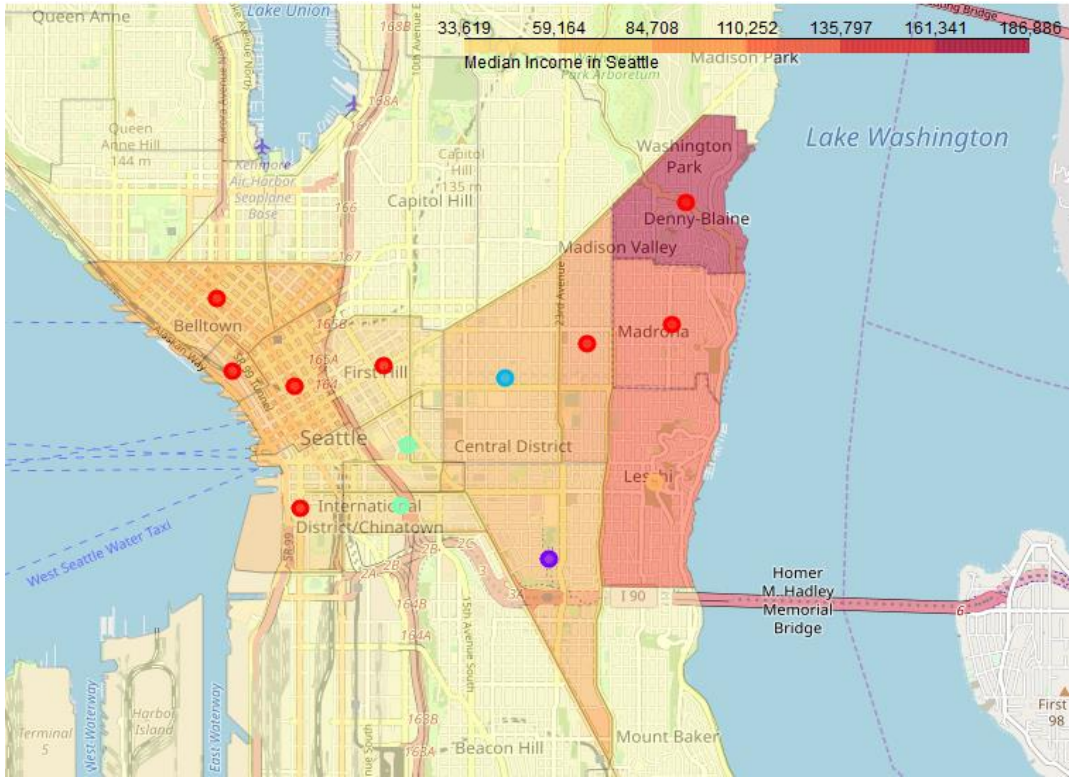
K-Means Clustering

- Set number of clusters to 5
- Examine most common venues in each cluster
- Determine similarities between neighborhoods in each cluster

Clusters

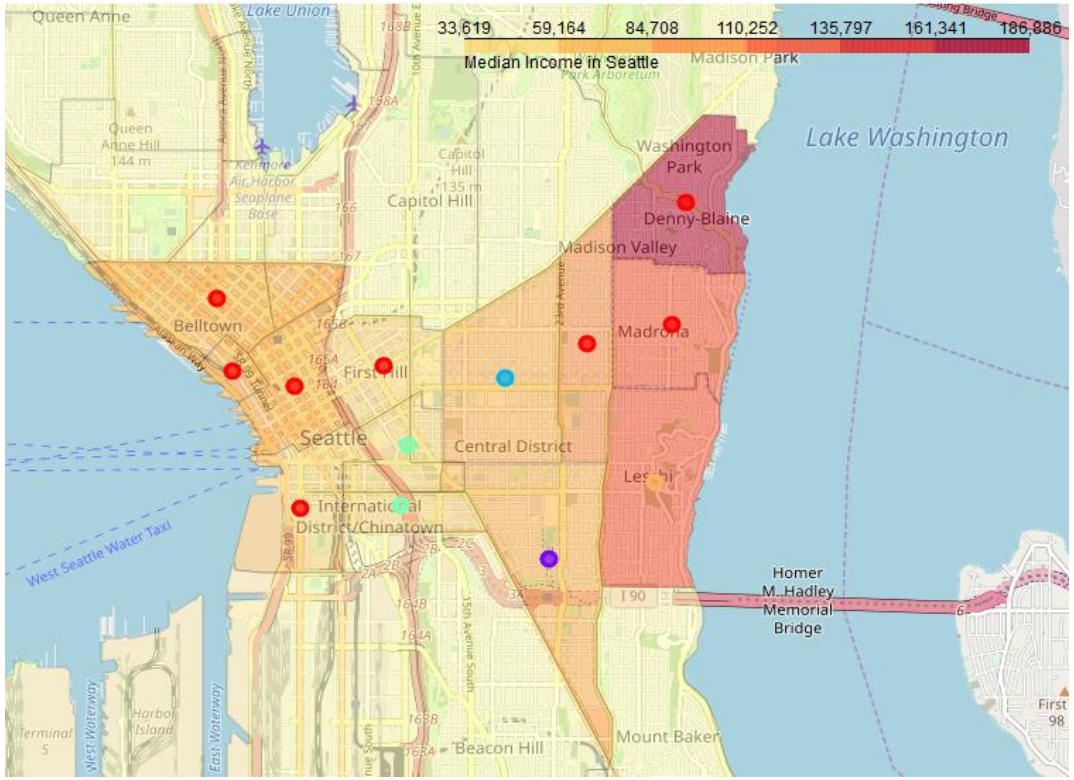
- Cluster 1:
  - Largest (8 neighborhoods)
  - Most common venues are coffee shops, but also several restaurants
- Cluster 2:
  - 1 neighborhood
  - Venues mostly related to travel and nature
- Cluster 3:
  - 1 neighborhood
  - Many entertainment venues: soccer stadium, theater, indie theater, art gallery
- Cluster 4:
  - 2 neighborhoods
  - Mainly Asian (Chinese, Vietnamese, Korean, Thai, and Japanese) restaurants
- Cluster 5:
  - 1 neighborhood
  - Recreation (park, playground, yoga studio), some restaurants

Clusters and Median Income



- Largest cluster (red) also contains neighborhoods with highest median incomes
  - Can potentially set higher menu prices in those neighborhoods
  - Potentially good location for restaurants
  - Will need to assess types of restaurants to determine how much competition is in each neighborhood

Clusters and Median Income



- Second largest cluster (light green)
  - Many Asian restaurants, so may be a lot of competition
  - Low median income, so menu prices will need to be lower
- Three clusters with one neighborhood each (light blue, purple, and yellow-orange)
  - Not optimal locations
  - Few venues in each neighborhood

Conclusion

- Using data to determine optimal locations can be especially useful for future business owners and investors
- Foursquare API data can provide information on whether there will be competition
- Future analyses
  - Optimal k not distinct from elbow method, may try different number of clusters
  - May expand to other districts/neighborhoods
  - Take into account other factors (e.g., population, real estate availability)