Finding a Compatible Neighborhood

IBM Data Science Capstone

Introduction

- The target audience for this project is people who are newcomers or considering a move to Portland, Oregon.
- The purpose of the project is to provide the user with a list of neighborhoods that are most compatible with a list of preferences that they submit.
- These neighborhoods would be promising areas in which to begin a housing search.

Data Required

- First we need a list of all Portland neighborhoods.
- Next we need latitude and longitude location data for each neighborhood.
- Relevant venue information for each neighborhood will be provide by querying Foursquare.
- User preferences for neighborhood amenities will be entered by the user.

Neighborhood Data

- List of neighborhoods with locations

	OBJECTID	NAME	LAT	LONG
0	1	CATHEDRAL PARK	45.5875	-122.7625
1	2	UNIVERSITY PARK	45.5785	-122.7318
2	3	PIEDMONT	45.5651	-122.6682

	OBJECTID	NAME	LAT	LONG
96	97	BRIDGETON	45.6018	-122.6693
97	98	EAST COLUMBIA	45.5908	-122.6522
98	99	USER_DEFINED	45.4334	-122.4970

Methods

- Query Foursquare for each neighborhood
- Summarize Foursquare results
- Combine Foursquare results with user preferences
- Segment neighborhoods with clustering algorithm
- Return and map results

Foursquare Query

- Query Foursquare for each neighborhood
- Limit to 100 venues
- Search radius set to 600 meters

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CATHEDRAL PARK	45.5875	-122.7625	Cathedral Park	45.587744	-122.759822	Park
1	CATHEDRAL PARK	45.5875	-122.7625	Occidental Wursthaus	45.588864	-122.761344	German Restaurant
2	CATHEDRAL PARK	45.5875	-122.7625	Occidental Brewing Company	45.588807	-122.761680	Brewery
3	CATHEDRAL PARK	45.5875	-122.7625	Hoplandia Beer	45.589662	-122.755614	Beer Store
4	CATHEDRAL PARK	45.5875	-122.7625	Cathedral Park Restaurant	45.588915	-122.761391	Café

Summary Stats for Venues per Neighborhood

- Check number of venues in typical neighborhood

count	96.000000
mean	23.135417
std	24.558682
min	1.000000
25%	6.000000
50%	15.000000
75%	32.000000
max	100.000000

Venue Category Proportions per Neighborhood

- Create summary profiles for each neighborhood from Foursquare data

	Neighborhood	ATM	Accessories Store	Adult Boutique	American Restaurant	Amphitheater	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asia Restaura
13	BUCKMAN COMMUNITY ASSOCIATION	0.0	0.0	0.0	0.021739	0.0	0.0	0.0	0.0	0.021739	0.0	0.021739	0.0	0

User Provided Preferences

- Show list of allowable categories:

```
['Neighborhood',
 'ATM',
 'Accessories Store',
 'Adult Boutique',
 'American Restaurant',
 'Amphitheater',
 'Antique Shop',
 'Arcade',
 'Argentinian Restaurant',
 'Art Gallery',
 'Art Museum',
 'Arts & Crafts Store',
 'Arts & Entertainment',
 'Asian Restaurant',
 'Assisted Living',
 'Athletics & Sports',
 'Auto Workshop',
 'Automotive Shop',
 'BBQ Joint',
```

- Sample user ordered preference list

```
#Make a user-defined preference list based on top venue categories.
user_vc=['Bookstore','Coffee Shop','Café','Park','Ice Cream Shop','Hotel']
user_vc
```

```
['Bookstore', 'Coffee Shop', 'Café', 'Park', 'Ice Cream Shop', 'Hotel']
```

User-Defined Neighborhood

- Store user's preferences as weights for each category in dataframe

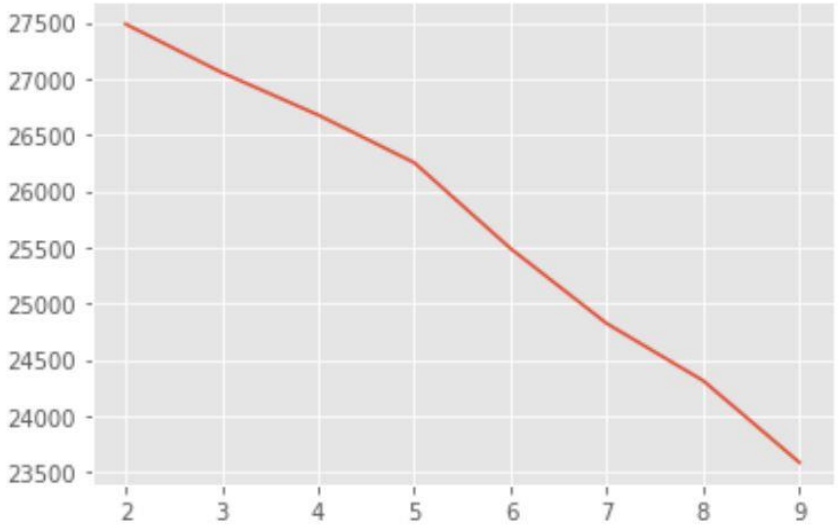
	Neighborhood	ATM	Accessories Store	Adult Boutique	American Restaurant	Amphitheater	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant
0	USER_DEFINED	0	0	0	0	0	0	0	0	0	0	0	0	0

- Append to venue proportions dataframe

	Neighborhood	ATM	Accessories Store	Adult Boutique	American Restaurant	Amphitheater	Antique Shop	Arcade	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Resta
92	WILKES COMMUNITY GROUP	0	0	0	0	0	0	0	0	0	0	0	0	
93	WOODLAND PARK	0	0	0	0.03125	0	0	0.03125	0	0	0	0	0	
94	WOODLAWN	0	0	0	0	0	0	0	0	0	0	0	0	
95	WOODSTOCK	0.0243902	0	0	0	0	0	0	0	0	0	0	0	0.024
96	USER_DEFINED	0	0	0	0	0	0	0	0	0	0	0	0	

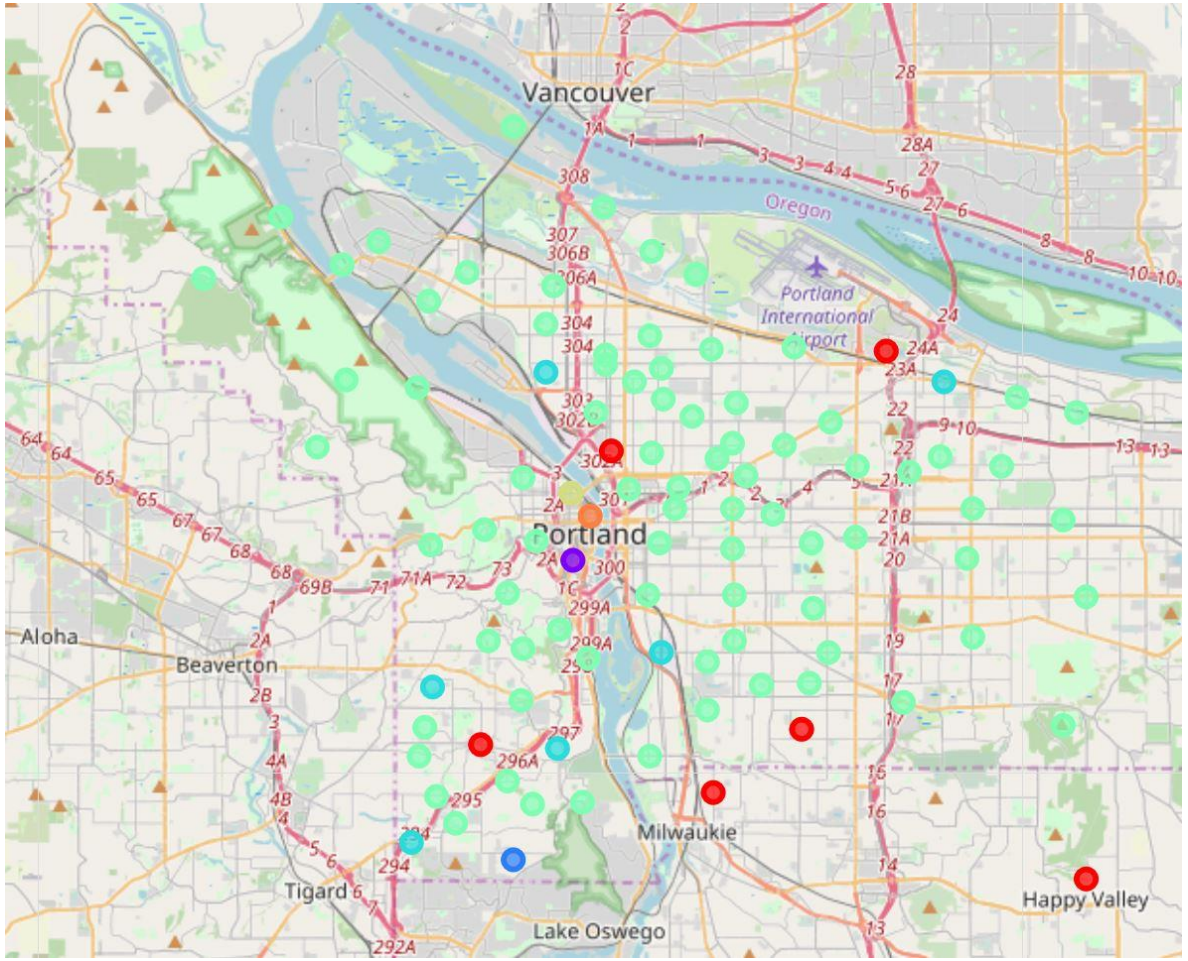
Segmenting Neighborhoods

- Use K-means clustering
- Select K=7 based on inertia
- Note “elbow” at K=7
- Standardize data before sending to the algorithm



Results

- Mapping Clusters
- User's neighborhood is represented by marker in southeast corner



Results

- Partial list of neighborhoods in user neighborhood's cluster

	NAME	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
9	SUMNER ASSOCIATION OF NEIGHBORS	Coffee Shop	Hotel	Business Service	Furniture / Home Store	Eye Doctor	Electronics Store	Elementary School	Ethiopian Restaurant	Event Service	Event Space
41	MULTNOMAH	Coffee Shop	Café	Music Store	Sports Bar	Thai Restaurant	Bookstore	Bar	Jewelry Store	Bakery	Toy / Game Store
42	BRENTWOOD-DARLINGTON	Trail	Dog Run	Park	Ice Cream Shop	Deli / Bodega	Night Market	Zoo Exhibit	Event Space	Elementary School	Ethiopian Restaurant
59	ARDENWALD-JOHNSON CREEK	Furniture / Home Store	Park	Café	Grocery Store	Coffee Shop	Dry Cleaner	Electronics Store	Elementary School	Ethiopian Restaurant	Event Service
79	ELIOT	Brewery	Dive Bar	Lounge	Coffee Shop	Sporting Goods Shop	Bookstore	Café	Ethiopian Restaurant	Tapas Restaurant	Nightclub
98	USER_DEFINED	Bookstore	Coffee Shop	Café	Park	Ice Cream Shop	Hotel	Eye Doctor	Elementary School	Ethiopian Restaurant	Event Service

Discussion

- One very large neighborhood is a frequent occurrence
- Limits usefulness of results if user’s neighborhood is in very large cluster
- User’s neighborhood occasionally in its own cluster
- Based on top ten venues, some neighborhoods in the same cluster appear to be very dissimilar

Future Work

- Different parameter choices could produce better results:
  - Search radius
  - Search limit
  - Length/content of preference list
  - Weights based on preferences
- Likely most important factor affecting results is large number of unique categories (293)
- Consider ways of consolidating unique venue categories