IBM Capstone

IBM Data Science Professional Course Series

Ideal Neighborhoods for
Chicago Food Lovers

Blake Rocheleau
July 2020

Business Problem

- Moving to a big city like Chicago can be intimidating
- Many different attractions, such as parks, gyms, schools, malls, etc.
- Let’s emphasize food above all:
  - Where would be the best place in Chicago to move?
  - Food is the #1 priority!

Data

- Need
  - List of Chicago neighborhoods
  - Location information for each neighborhood
  - Venue information for each

Example of Data

	Neighborhood	Latitude	Longitude
0	Albany Park	41.96829	-87.72338
1	Altgeld Gardens	41.65441	-87.60225
2	Andersonville	41.98046	-87.66834
3	Archer Heights	41.81154	-87.72556
4	Armour Square	41.83458	-87.63189
5	Ashburn	41.74785	-87.70995
6	Ashburn Estates	41.74785	-87.70995
7	Auburn Gresham	41.74319	-87.65504
8	Avalon Park	41.74507	-87.58816
9	Avondale	41.93925	-87.71125

Methodology

- Scraped data from Wikipedia (Beautiful Soup Python library)
  - Store date in a Pandas dataframe
- Obtain coordinates of each Chicago neighborhood (Geocoder)
- Obtain venues for each neighborhood (Foursquare API)
- Use one-hot encoding to prepare data for clustering
- Apply k-means clustering
- Analyze results from k-means

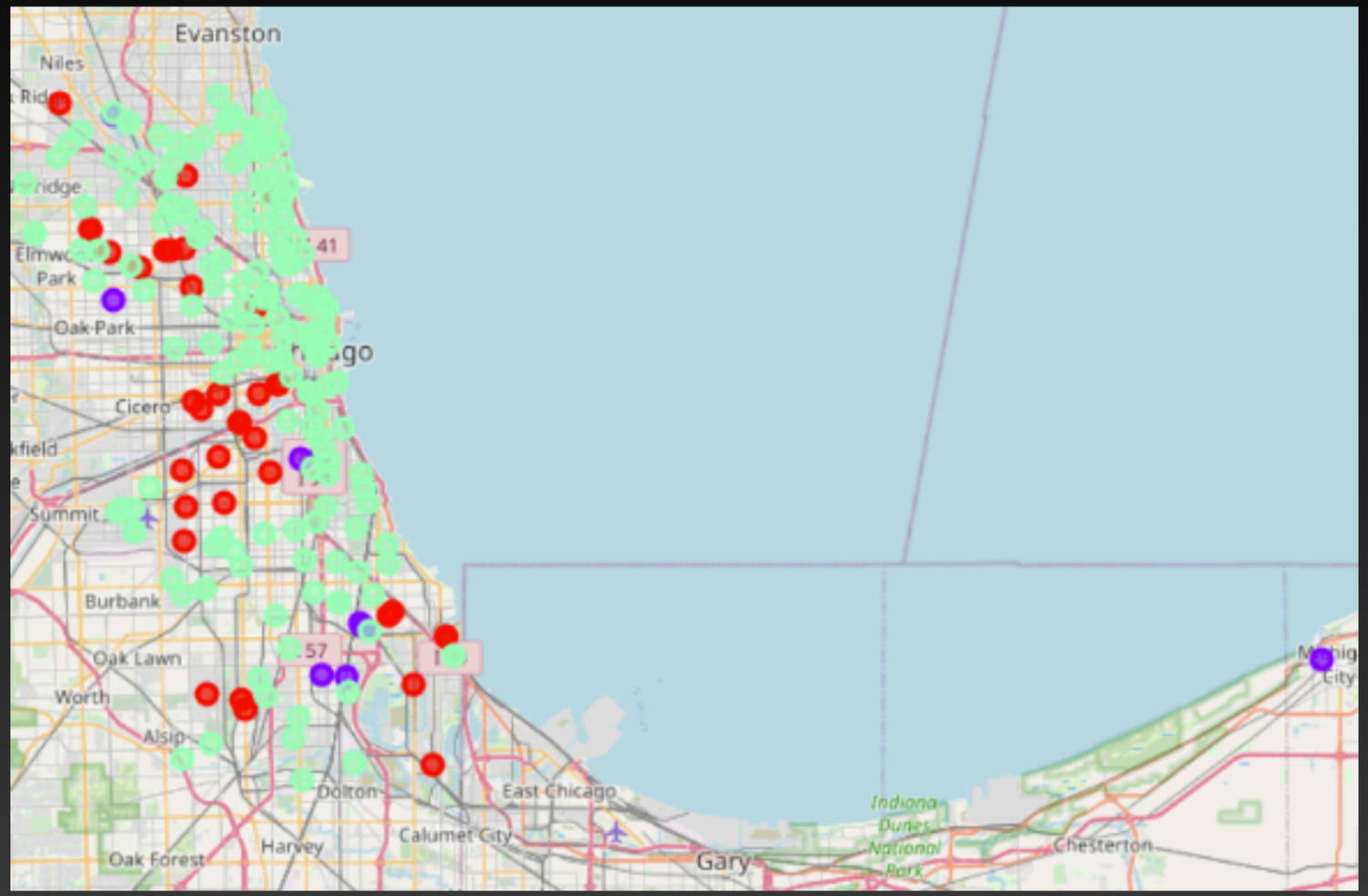
Example Foursquare API call

```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&cli
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    search_query,
    lat,
    lng,
    radius,
    LIMIT)

# make the GET request
results = requests.get(url).json()["response"]['groups'][0]['items']
```

Results

K-means Clustering Results



Cluster Analysis

- 3 resulting clusters:
  - Cluster 0: neighborhoods w/ low number of food options
  - Cluster 1: neighborhoods w/moderate number of food options
  - Cluster 2: neighborhoods w/high number of food options

Discussion

- Neighborhoods closer to water (Lake Michigan) have more food options
- Cluster 2 (high food-density) contains a large number of neighborhoods
  - Plenty of options for food lovers!
- Further investigation involves adjusting radius in API call & considering other variables such as income and population density

Conclusion

- Answer to business problem: choose a neighborhood in cluster 2 when considering best location for food lovers
- Findings will allow food lovers moving to Chicago to find a location that will make them happy