Neighborhood Similarities in Different Areas

Xiaojun Fu

Target users (why neighborhood similarities?)

- People moving to new place:
  - ➢For a smooth transfer in a new place

- Business expansion (restaurants, shopping malls,…)
  - ➢Learn about the neighborhood and find the right location

Preliminary data analysis

- Goal
  - ➤ Find similar cities in Minnesota, US and Tianjin, China

- Data resources
  - ➤ City names: Wiki
  - ➤ Latitudes and longitudes: Geopy library
  - ➤ Neighborhood venues: Foursquare API

- Methodology
  - ➤ K-nearest neighbors clustering (Kmeans package) on top 20 frequent venue categories for each area

Data exploration:

- Top frequent venue categories

City	State	Country	Latitude	Longitude	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
Hedong	Tianjin	China	39.118836	117.240799	Shopping Mall	Multiplex	Sporting Goods Shop	Hotpot Restaurant	Asian Restaurant	Fast Food Restaurant	Coffee Shop
Hexi	Tianjin	China	39.108506	117.217056	Coffee Shop	Fast Food Restaurant	Hotel	Metro Station	Park	Sandwich Place	Café
Hebei	Tianjin	China	39.161741	117.213204	Supermarket	Fast Food Restaurant	Metro Station	Park	Train Station	Electronics Store	Elementary School
Hongqiao	Tianjin	China	39.164545	117.146467	Fast Food Restaurant	Art Museum	Bakery	Chinese Restaurant	Hotel	Garden Center	Garden
Eden Prairie	Minnesota	United States	44.854686	-93.470786	Coffee Shop	Insurance Office	Electronics Store	Wine Bar	Non-Profit	Skating Rink	Park
Edina	Minnesota	United States	44.889703	-93.350122	Art Museum	Burger Joint	Soccer Field	Football Stadium	Gym / Fitness Center	Bar	Park

Neighborhood similarity results

- K-nearest neighbors clustering
  - KMeans package from sklearn.cluster

- Hyperparameter
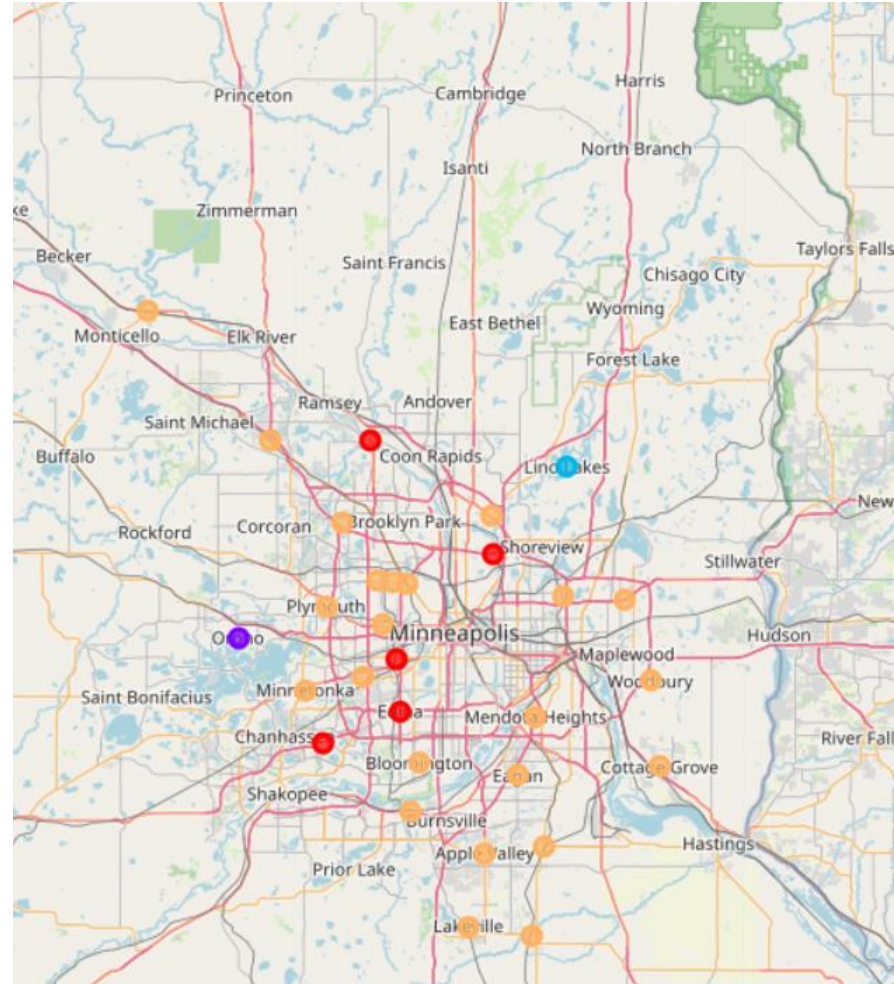 - Cluster # = 5
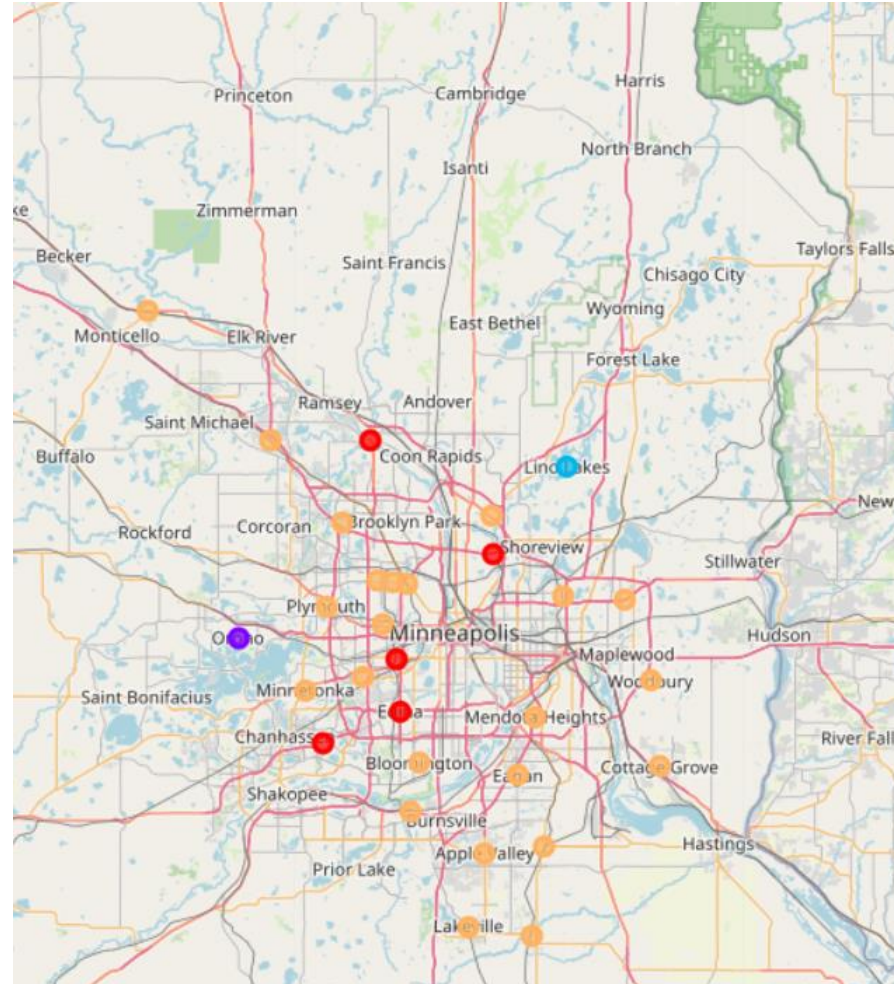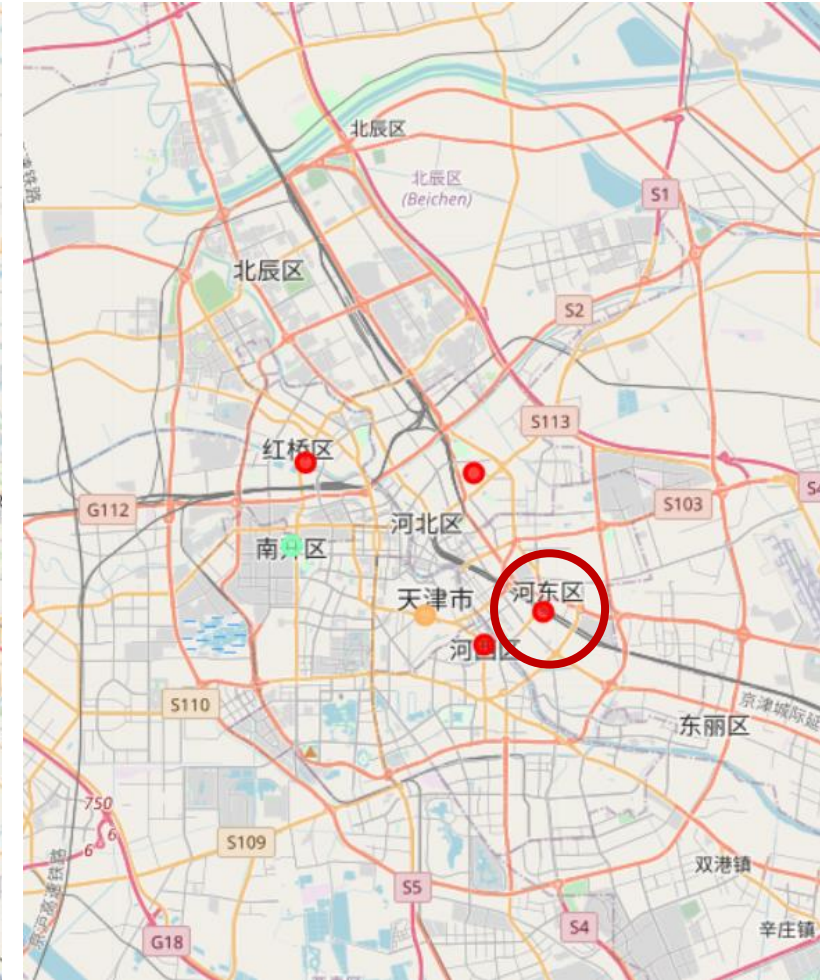

Minnesota, US
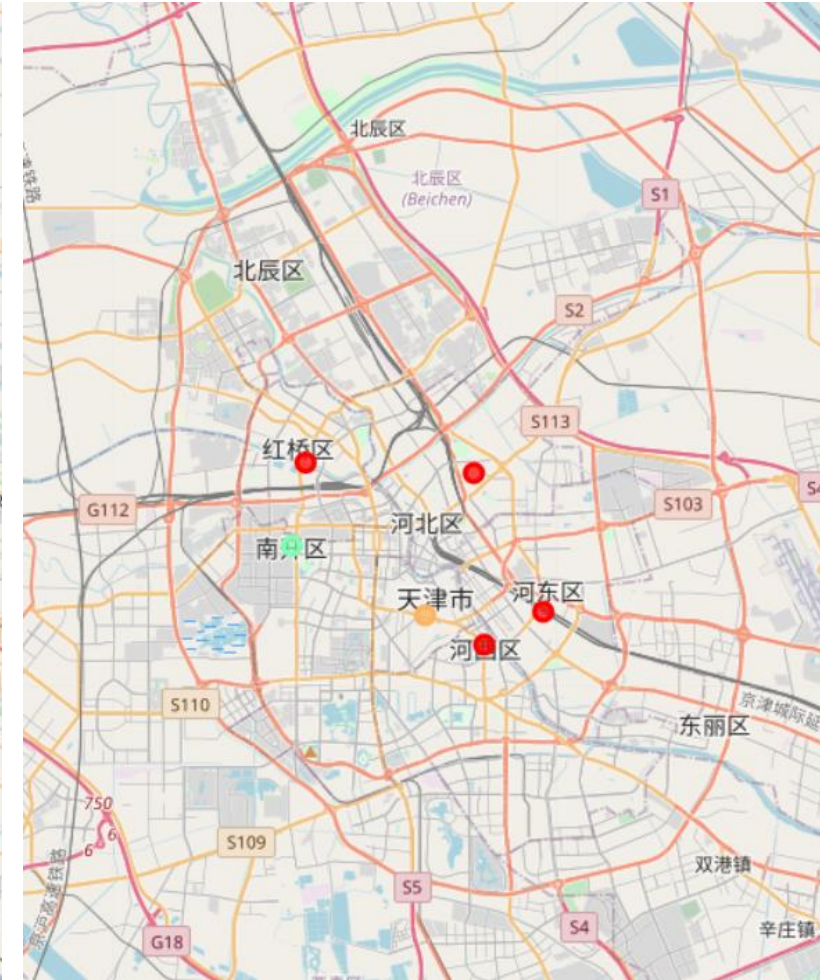
Tianjin, China

If I am the target user…

- My hometown: ⭕

- Candidates: red places

- Further investigations:
  - ➢ Red places: featured with Chinese culture, e.g. Chinese school & groceries
  - ➢ Validate the results

Minnesota, US

Tianjin, China

Future work

- Tune hyperparameter
  - ➢Choice of cluster numbers

- Try other methods without setting cluster numbers
  - ➢Hierarchical agglomerative clustering method

- Analysis for more neighborhoods
  - ➢International and within US

Thanks!