Clustering analysis for branch location (hypothetical)

IBM Applied Data Science Capstone project

Emil C. Fischer, May 2020

Introduction

Finding a new Toronto location for an established NYC branch

- The Williamsburg location sells more cups than in the Upper East Side
- The chain thinks the local business environment is involved

Example used in this project

A hypothetical—but realistic—scenario

- An NYC coffee shop chain wants to open a new shop in Toronto
- Their store in Williamsburg, Brooklyn sells more cup than their store in the Upper East Side, Manhattan.
- They think this is due to the local business environment.

Question: What **neighborhood in Toronto** is most like Williamsburg, Brooklyn?

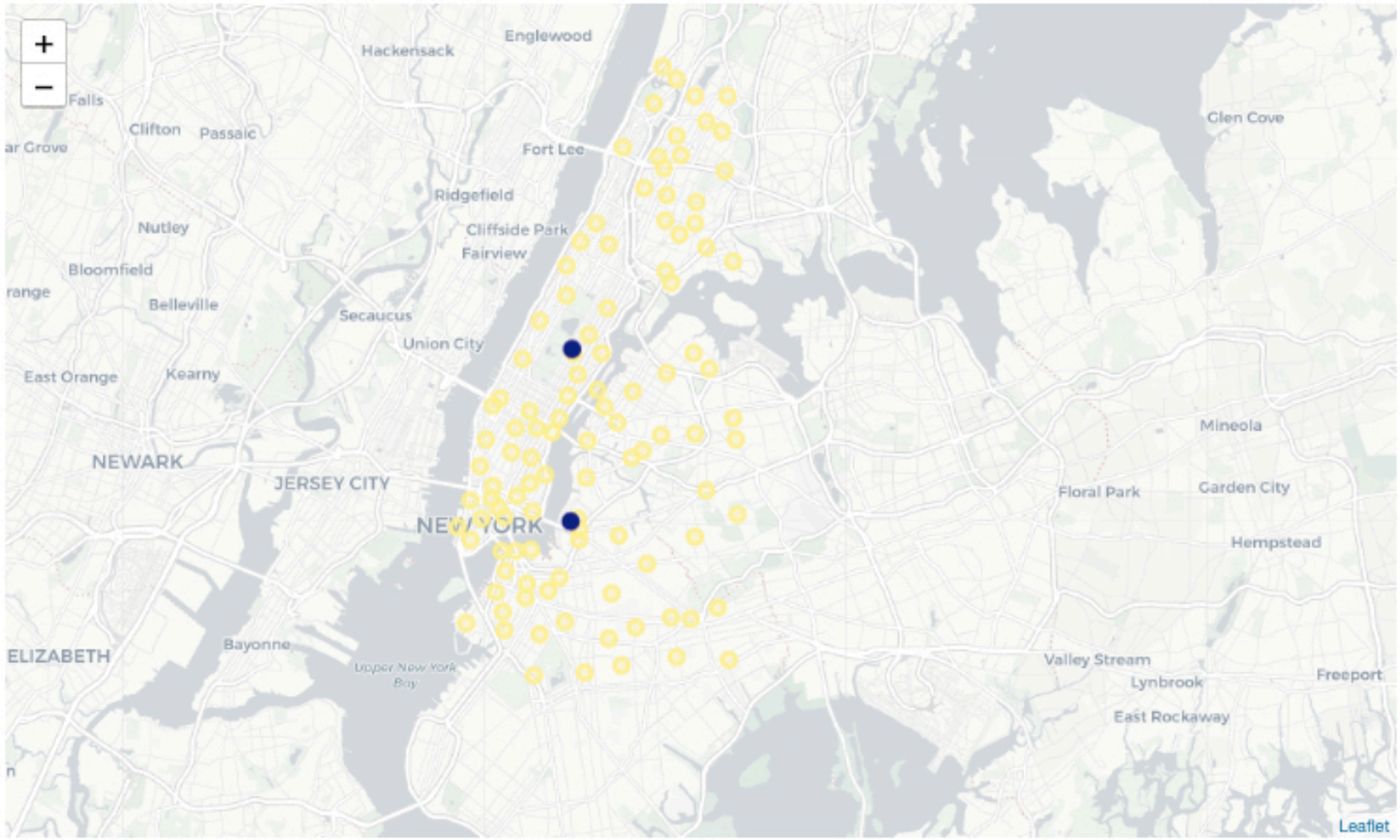
Data

The Foursquare API can access local business data

- provide data on nearby businesses based in latitude / longitude
  - including categories

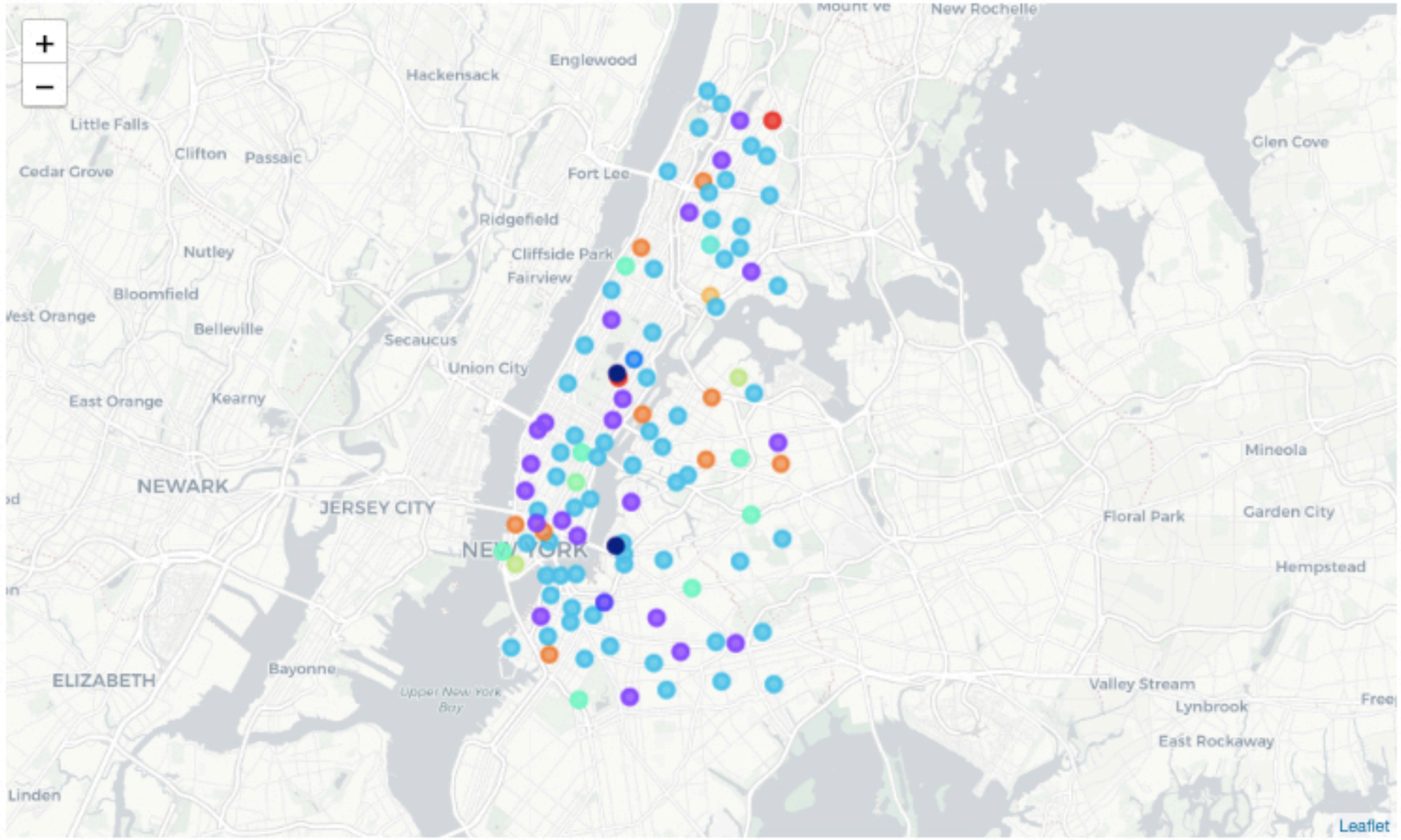
Neighborhoods in NYC

Locations included (yellow) and coffee shops (navy blue)



Clusters in NYC

Color coded (k = 12) clusters and coffee shops (navy blue)

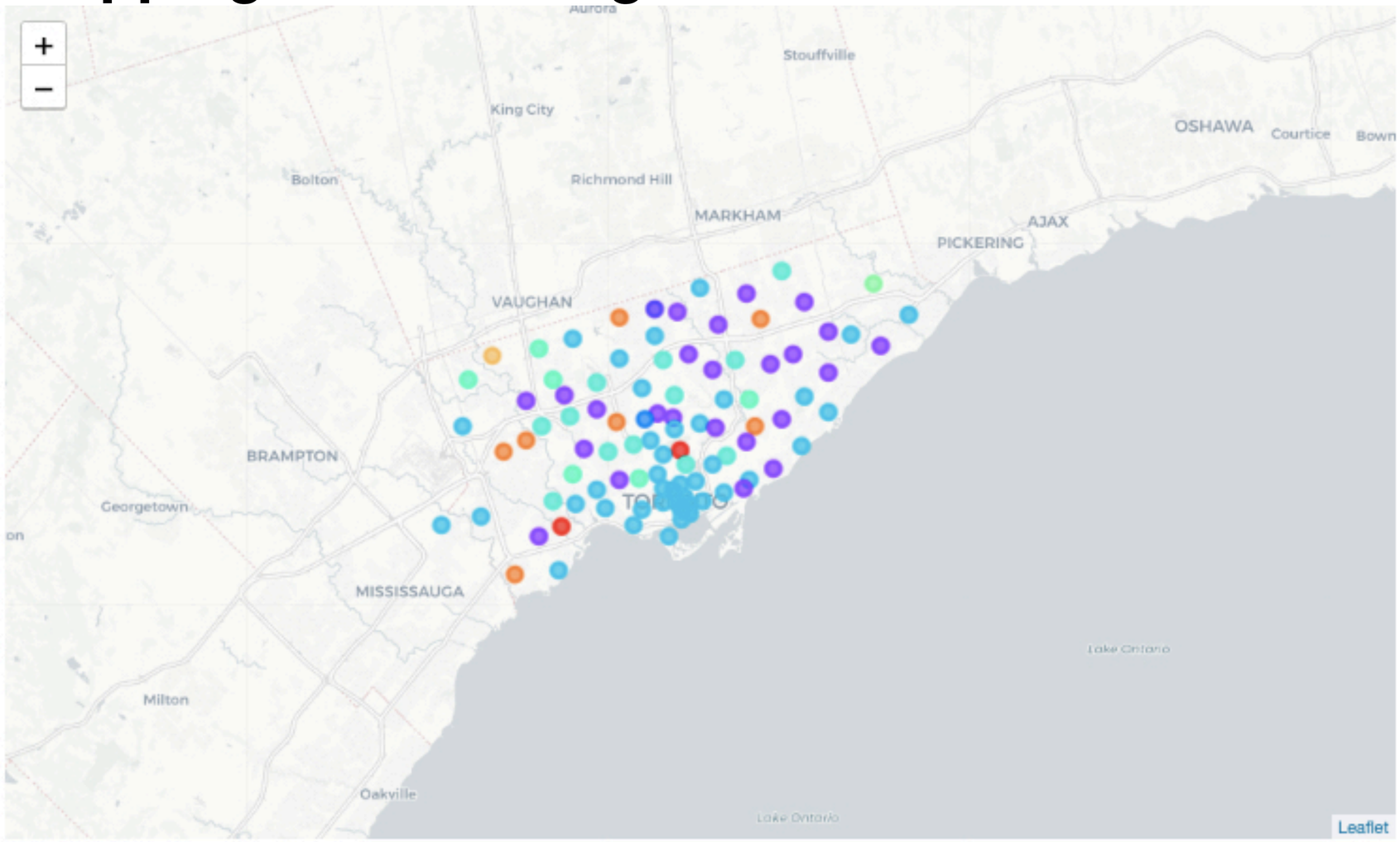


Upper east side: cluster 11

Williamsburg: cluster 4

Clusters in Toronto

Mapping Toronto neighborhoods to NYC cluster centers



45 members of cluster 4

Making a better model

More information can allow for a more accurate model

- Profile on shoppers in each neighborhood
- Popularity and visitor information for nearby venues
- Brand value of nearby venues

Conclusion

Finding a new Toronto location for an established NYC branch

- 45 neighborhoods (postal codes) have a similar composition of businesses to that of Williamsburg, Brooklyn
- Hence… Toronto has plenty of potential locations for a new coffee shop
- More data can provide a more accurate (and more predictive) model