Using Foursquare location data to determine clusters for potential new business locations

Alvin Jin

IBM Data Science Capstone

Introduction/Business Problem

- Restaurants and bars have been closing in Toronto in 2020 due to the COVID-19 pandemic[1].

- Some businesses, like those related to the pet industry have seen strong gains in the same time period[2].

- Together, these factors suggest an opportunity for a bar and dog park once restrictions are lifted.

- Businesses in the Canadian service sector see greater turnover than any other industry[3].

- Identifying proper location is essential for any new business to maximize success.

1. https://nowtoronto.com/food-and-drink/toronto-restaurants-closed-2020-during-covid-19
2. https://www.retail-insider.com/retail-insider/2021/01/pet-focused-retailers-in-canada-see-sales-gains-during-pandemic-amid-work-from-home/
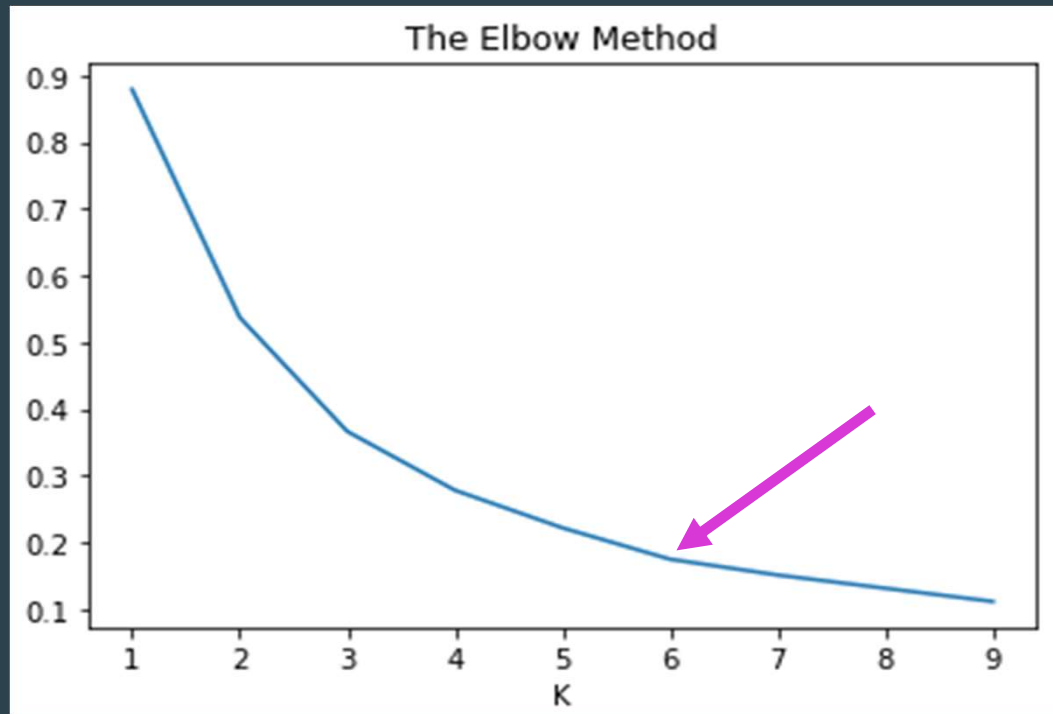3. https://www.ic.gc.ca/eic/site/061.nsf/eng/h_03126.html

Data

- Postal codes and neighborhoods scraped from Wikipedia

- Venue name, category, and geographic location was extracted from Foursquare
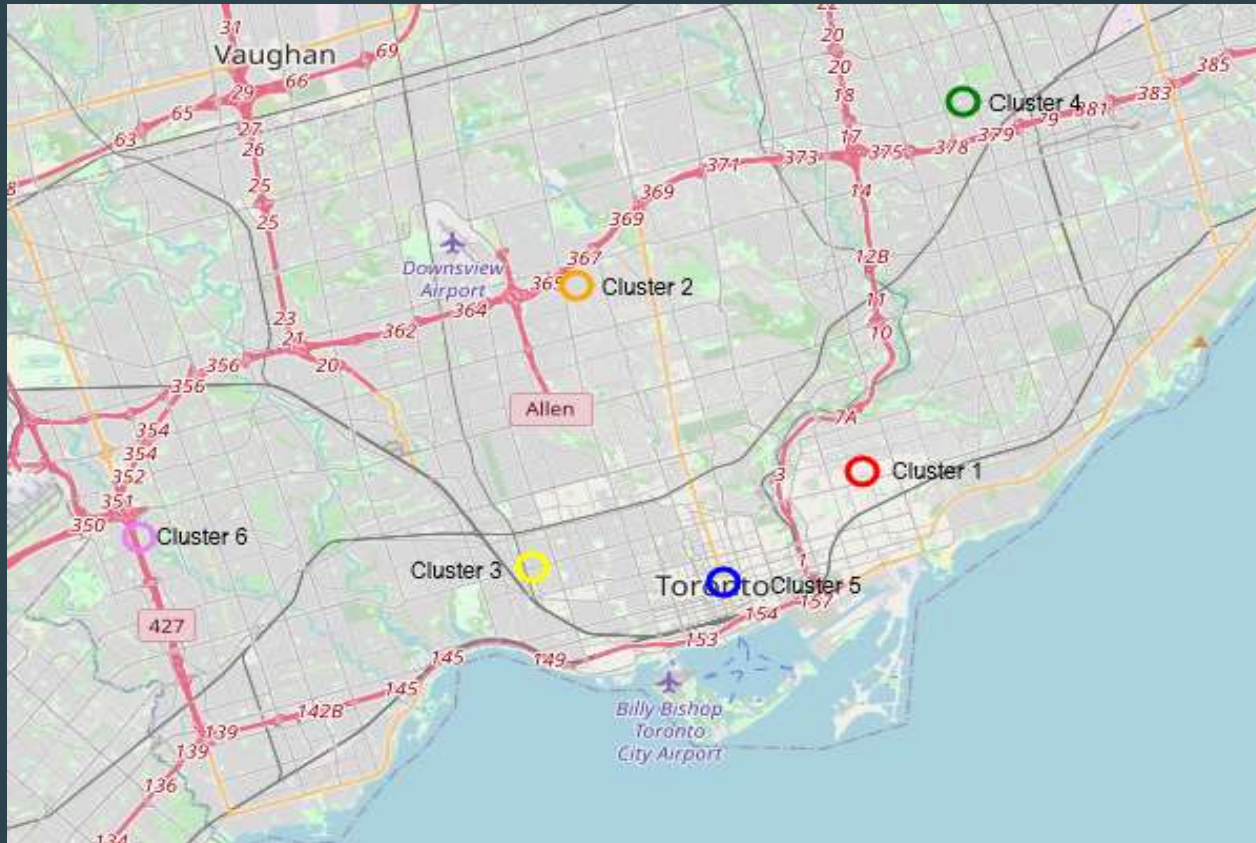
Methodology

- Venue data limited to categories related to drinking, pets, and outdoor activity.
  - 215 total venues
- Elbow method used to determine optimal number of clusters.
- K-means clustering to identify potential new business locations

The Elbow Method for Optimal Number of Clusters

Graph appears to flatten after a K of six

Toronto Cluster Centers

Discussion

- Six potential centers were identified around Toronto.

- These can be used as starting points for searching for a good business location and future expansion opportunities.

- The model could be refined more.
  - Different weights could be added to the venues.
  - Add income and wealth into calculations.
  - Add pet ownership and online pet supplies sales data.
  - Include foot traffic through areas.

Conclusion

- It is difficult to start a small business and many fail.

- Using machine learning to identify potential areas to start can help maximize chances of success.