



# PROFILING MAJOR CITIES BASED ON VENUES USING CLUSTERING

RIFAT JABBAR



# CITY PROFILING

## I n t r o d u c t i o n

- Compare major urban centres - Toronto and Montreal
- Perform clustering based on venue types
- Assess characteristics of clusters

# DATA AND SOURCES

- Toronto postal codes – Wikipedia
- Montreal postal codes – Wikipedia
- Toronto postal code geographic coordinates – hosted in .csv file
- Montreal postal code geographic coordinates – hosted in .csv file
- Venue data (name, geographic coordinates, type) from Foursquare API

# METHODOLOGY

1

- Scrape postal code data from Wikipedia
- Append geographic coordinates for each neighbourhood

2

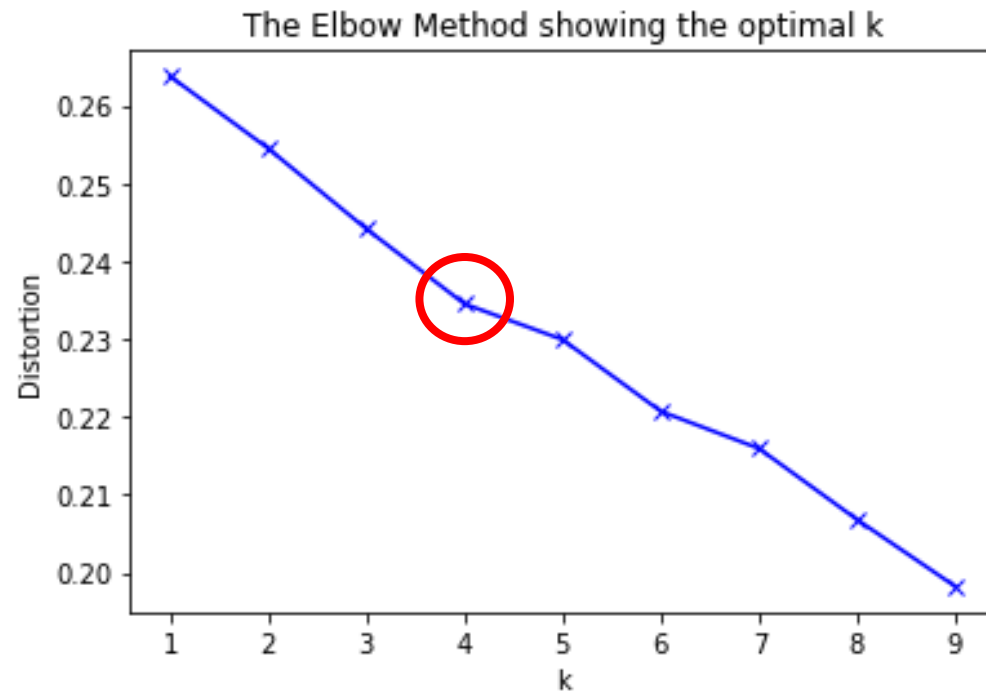
- Gather venue data from Foursquare API
- Determine top ten venue types for each neighbourhood

3

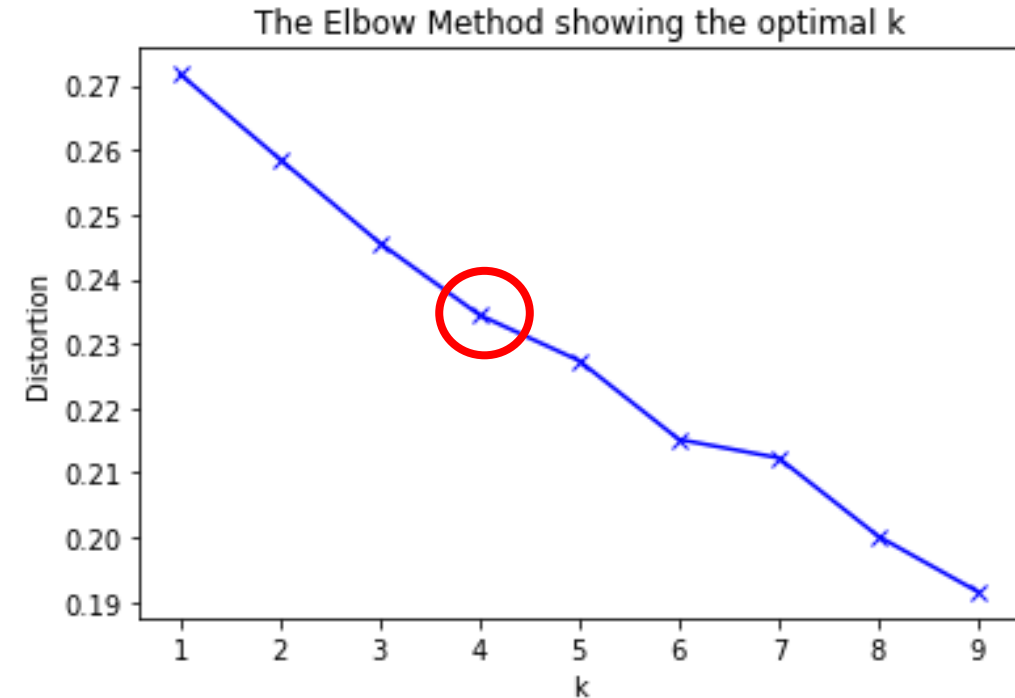
- Use elbow method to determine optimum k
- Perform k-means clustering

# RESULTS

- Optimum k for k-means clustering



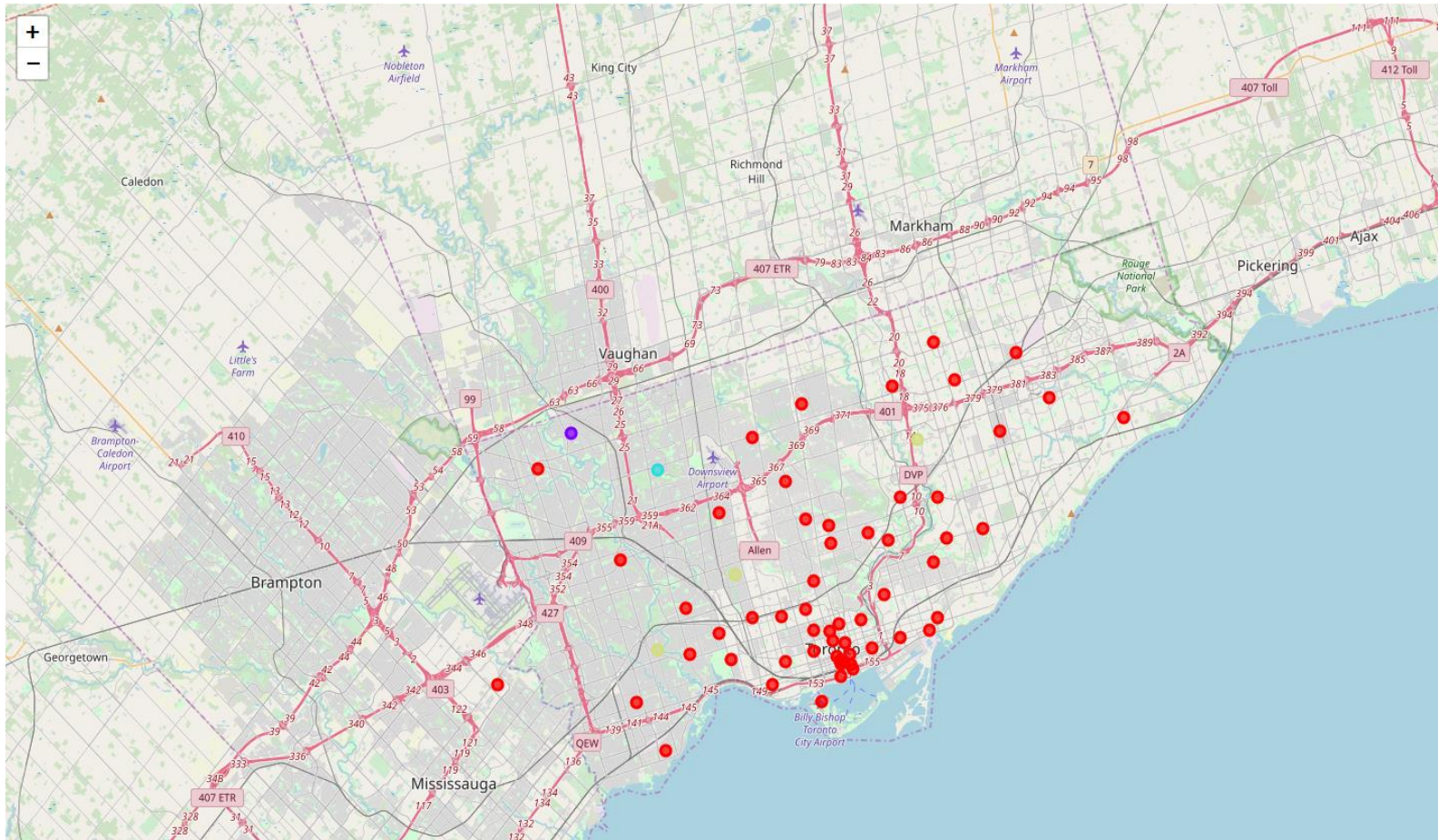
Toronto



Montreal

# RESULTS

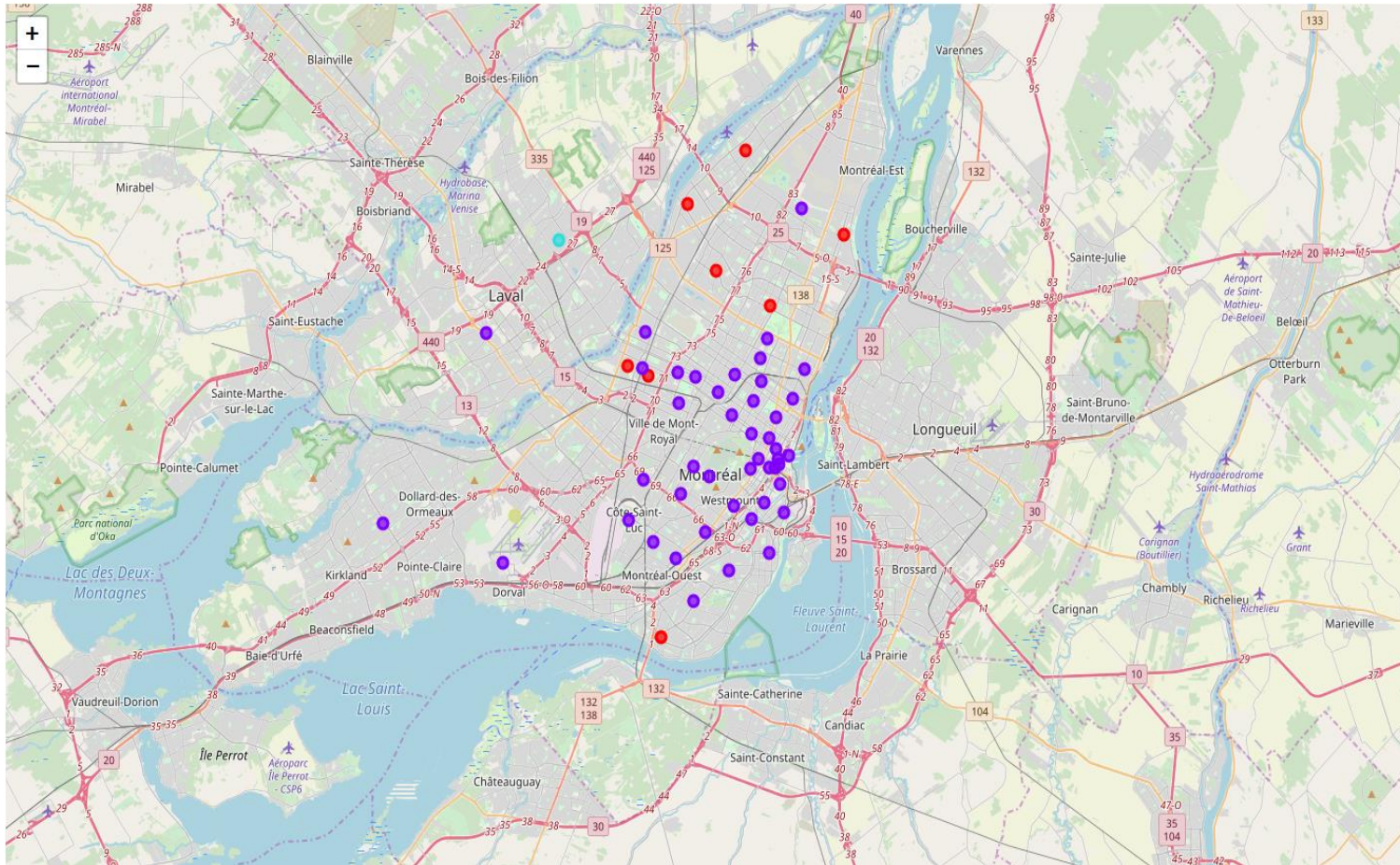
- Toronto clusters – mapped using Folium package





# RESULTS

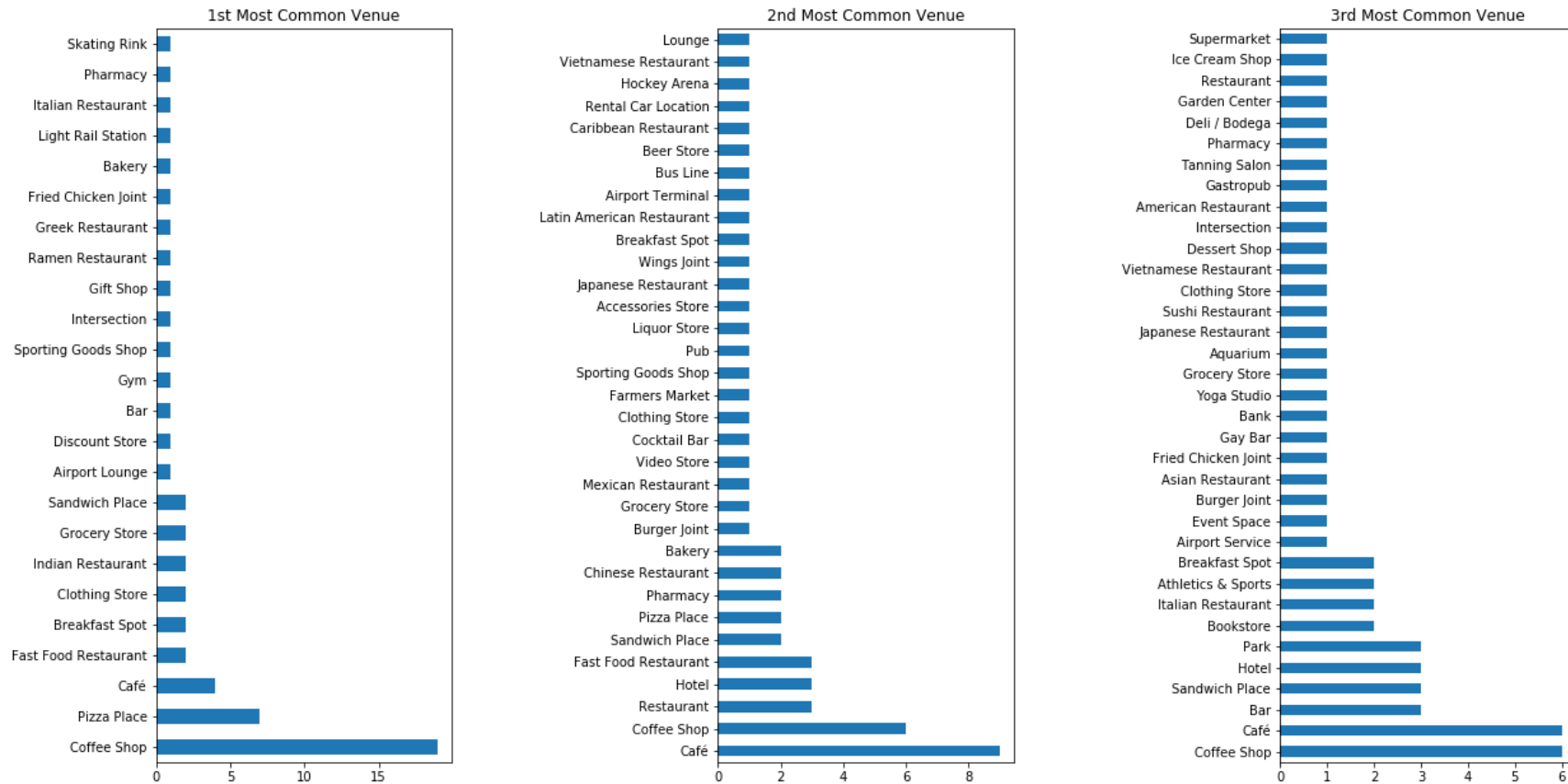
- Montreal clusters – mapped using Folium package



# RESULTS

- Toronto – most common venues in largest cluster

Venues in largest cluster

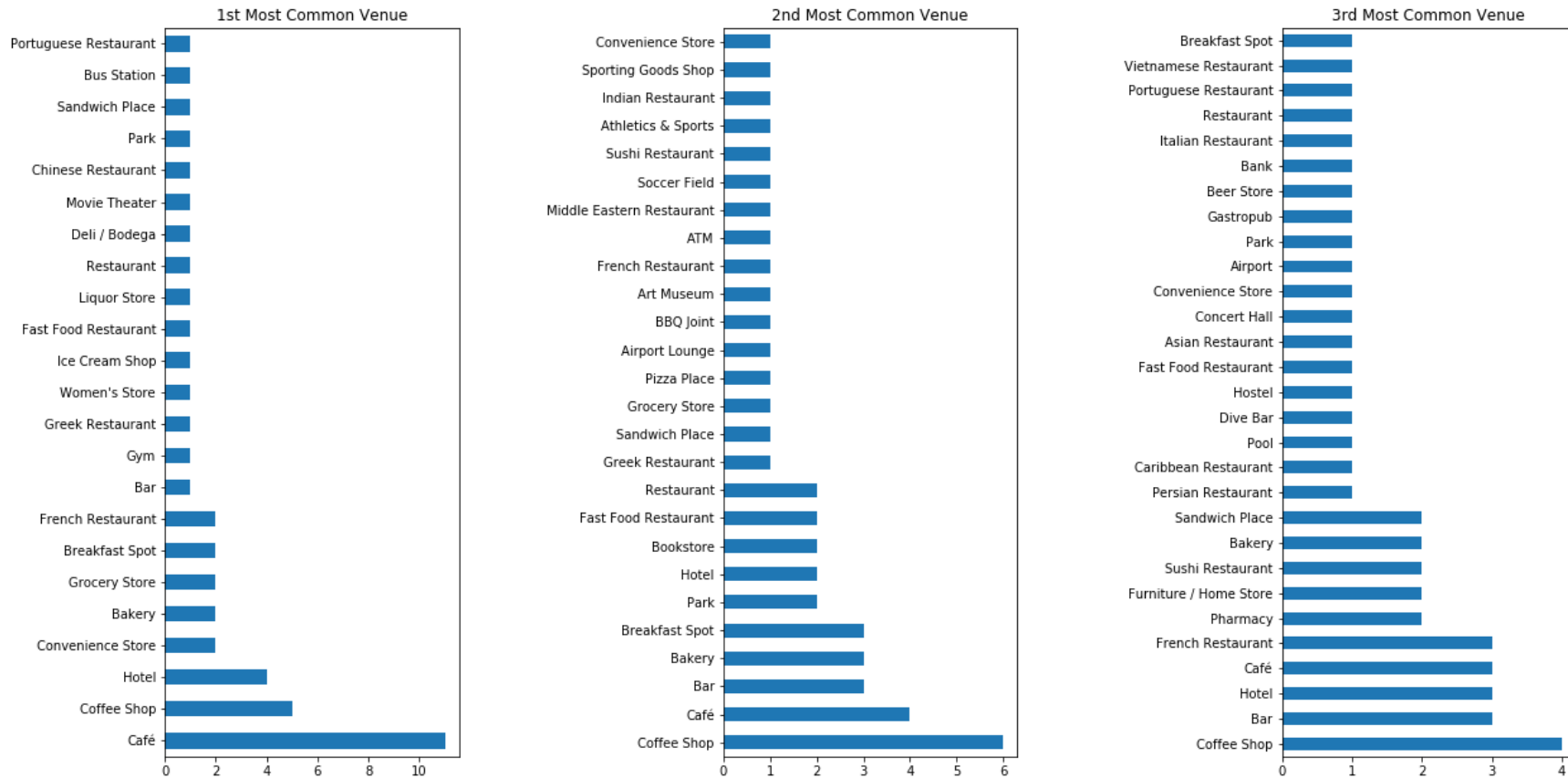




# RESULTS

- Montreal – most common venues in largest cluster

Venues in largest cluster



# OBSERVATIONS

- Both cities exhibit many similarities in their largest cluster
- Most common venues include coffee shops, cafés, bars, hotels, and restaurants
- This clustering method can be used to profile similar major cities in North America
  - Quantities of different venue categories are indicative of competitive landscape
  - Can be used by businesses that would like to better understand a locale before launching an operation there



**THANK YOU**