

IBM Data Science Capstone

Opening a New Chinese Restaurant in Toronto

Introduction

- Chinese restaurants are an indicator of Chinese cultural influence in many cities. Oftentimes, they are isolated to certain distinct districts or Chinatowns.
- This project seeks to determine if there is a pattern in the distribution of Chinese restaurants in the city of Toronto and where would be an optimal location to open a new restaurant serving Chinese cuisine. The optimal location would consider factors that influence a food business such as proximity to a loyal customer base and competition from other restaurants.

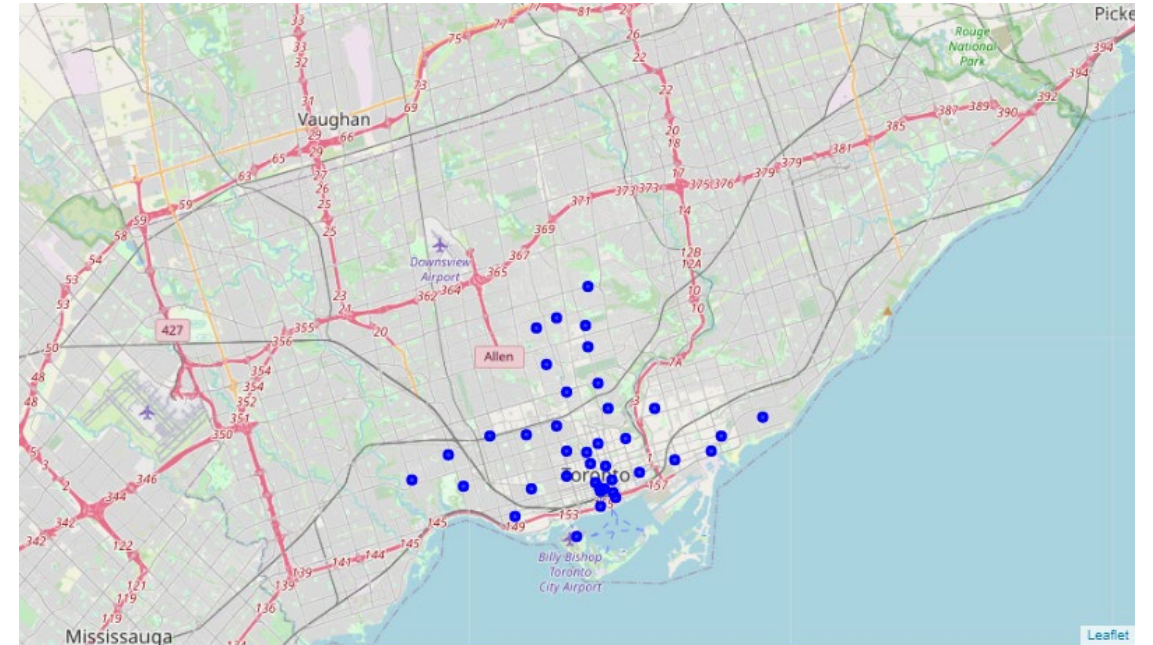
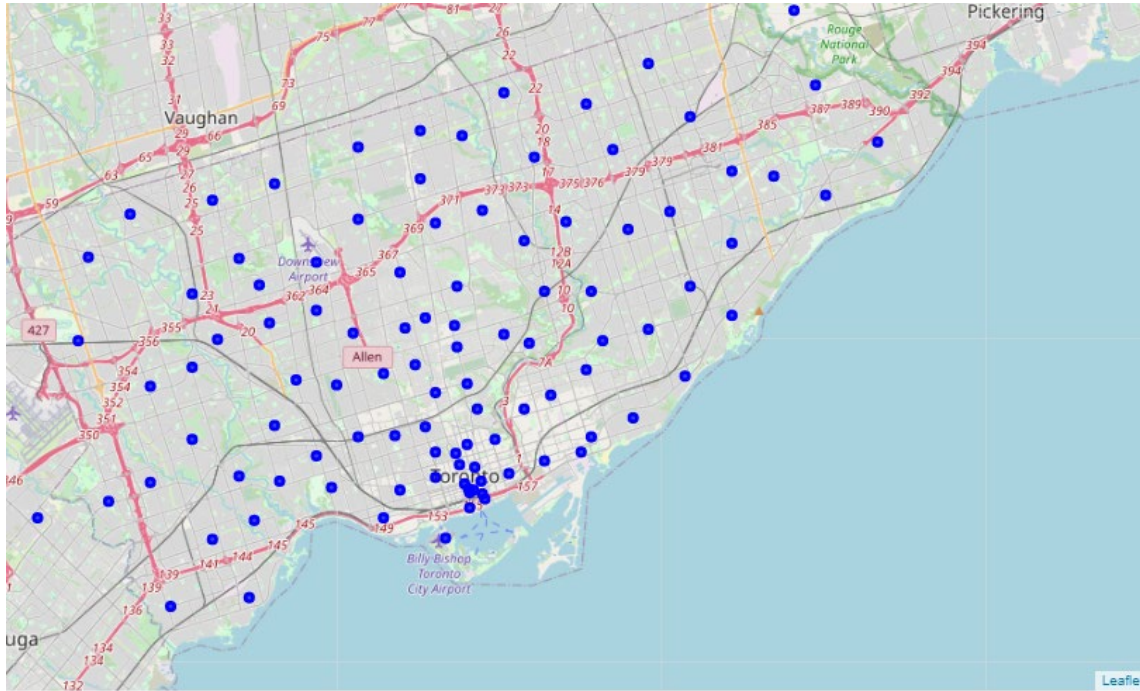
Data

- Wikipedia page containing list of neighborhoods of Toronto – Web scrapping technique to extract data from Wikipedia page, Python requests and BeautifulSoup packages were used.
- Geographical Coordinates of these neighborhoods – CSV file provided by Coursera (Geospatial_Coordinates.csv)
- Venues data for these neighborhoods – Foursquare API
- Map of Toronto – Python Folium package

Methods

- We get the list of neighborhoods in Toronto from Wikipedia page
- We get the geographical coordinates for those positions
- We get the venues data from the locations
- We prepare the data for K-means clustering
- We evaluate the clusters to determine where the best location to open a new Chinese restaurant is

Limiting to Neighborhoods with “Toronto”



Results



The highest concentration of Chinese restaurants is in cluster 1, downtown Toronto.

There are locations Adelaide, King, Stn A PO Boxes, Church and Wellesley

Conclusion

- We can open Chinese restaurants in clusters 1 and 2, in the following locations:

- Cluster 1

- Adelaide, King, Stn A PO Boxes, Church and Wellesley

- Cluster 2

- Studio District, Summerhill West, Rathnelly, South Hill, Forest