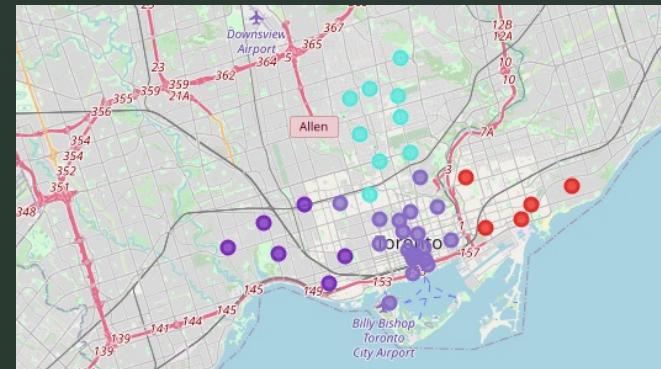


IBM Data Science Professional Capstone Project

The Battle of the Neighborhoods

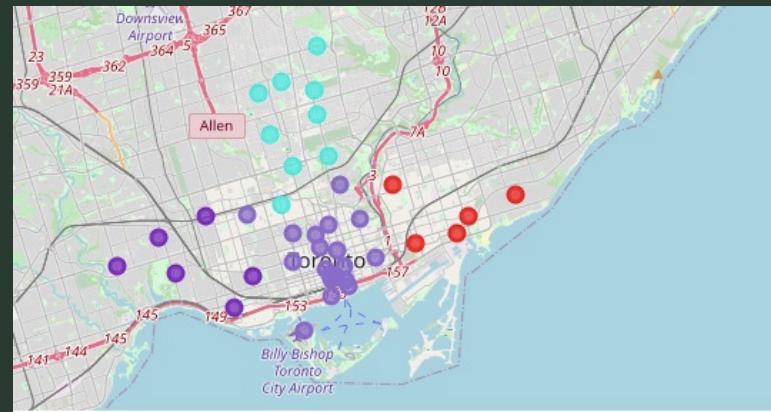
Introduction and Business Problem

- Clustering the Neighborhoods of Toronto
- Goal: To recommend a suitable location for the first restaurant in Toronto for a global Sushi Restaurant Chain in the city of Toronto



Data

- The Datasets used are:
 - A list of Toronto's neighborhoods grouped by boroughs (Wikipedia)
 - Neighborhood centroid geolocation data (IBM)
 - Venue Data for Japanese (and Asian) Restaurants (Foursquare)



Methodology

- Extracting neighborhood data using the Beautiful Soup package and storing it as a pandas dataframe.
- Extracting and merging the geolocation data with the neighborhood dataframe.
- Using Foursquare API to gather Venue data for upto 100 locations within a radius of 500 meters from each neighborhood centroid.
- Grouped the venue data by frequency of occurrence of "Sushi Restaurants" to form the basis for the analysis

Results

K-means clustering to create 3 distinct clusters of Neighborhoods

1. Cluster0 (red): little or no Sushi Restaurants
2. Cluster1 (purple): no Sushi Restaurants
3. Cluster2: Moderate – high number of Sushi Restaurants



Evaluation and Recommendation

- 23 Neighborhoods in the city of Toronto with 1 or more Sushi Restaurants
- Major Concentrations of Sushi Restaurants in south central (downtown) Toronto
- Recommendation: There are 2 areas that we recommend:
 1. The pocket of area starting from Central Bay st. upto North Midtown – these areas do not have a high density of Sushi restaurants although being very close to the center of the city.
 2. India Bazaar: Known as a popular district for other (Non-Japanese) Asian Restaurants and attracts clients looking for pan-Asian cuisine.

Limitations and Suggestions for Further Research

This project primarily takes into account the presence and density of Sushi Restaurants in Toronto, however there are many more factors that can be additionally considered:

- Foot-traffic Data
- Population Density and Demographic Distribution
- Mean, median and standard deviations of income of residents in each neighborhood
- Consumer Survey Data to identify the specific demand in the market
- Location Rentals and Capital Leasing costs to justify the financial feasibility of opening the restaurant

To analyze the merits for each neighborhood with all the above data is not only a long but also an expensive process. However, utilizing the above data sources will lead to a much more informed and complete recommendation that will better suit the client.

Conclusion

In this project, we have:

- Defined a business problem for a client
- Collected and pre-processed data from a variety of online sources
- Performed k-means clustering to generate insights from the data
- Generated Insights and Recommendations for the client

Sources

- [https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:
M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada_M)
- http://cocl.us/Geospatial_data
- <http://developer.foursquare.com/docs>

www.github.com/ritchatterjee

Thank You