

# Cluster Postal Areas in Toronto Based on Food

—Capstone project of IBM Data Science Specialization

September 2019

# Clustering areas based on cuisines in Toronto is valuable for both residence and travels

- Toronto is the provincial capital of Ontario in Canada with large population and great number of travelers
- Toronto has a variety of different cuisines due to large number of immigrants with great diversity
- Both residence and travelers are interested in living in a place close to their favourite cuisines
- Clustering postal areas based on cuisines can help residence and travels to find areas with similar cuisines

# Data sources

In this project, the following dataset will be used:

- A list of postal codes with boroughs and neighbourhoods in Toronto at [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).
- Geographical coordinates of each postal codes can be found at [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data).
- Foursquare API is used to get the most common venues of specified postal code and the venue category hierarchy.

# Data cleaning

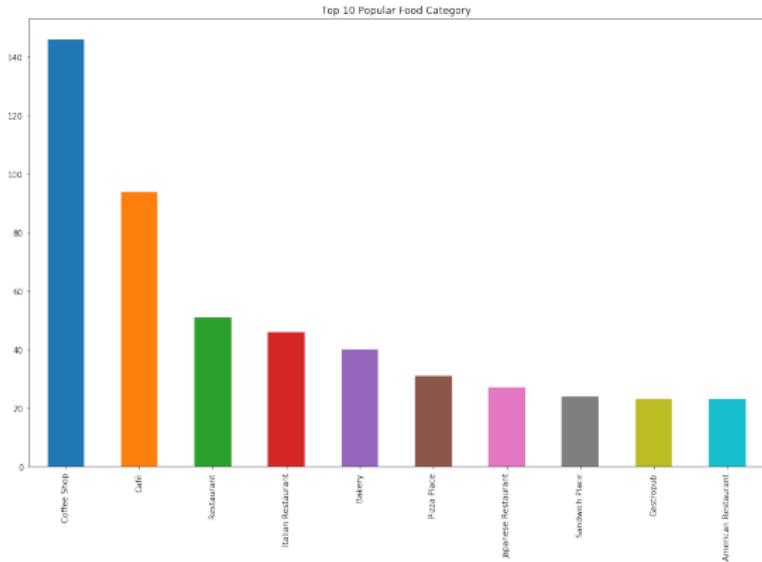
After data wrangling, geographical information and most common food venues of each postal code are found and used to cluster areas in Toronto and visualize on a Toronto map.

- Only postal areas in boroughs with word 'Toronto' in name are considered
- Only venues in Food category are considered
- Venue categories are converted to features using one hot encoding
- In total, 38 postal areas and 88 features in the cleaned dataset

# Postal areas superimposed on map of Toronto

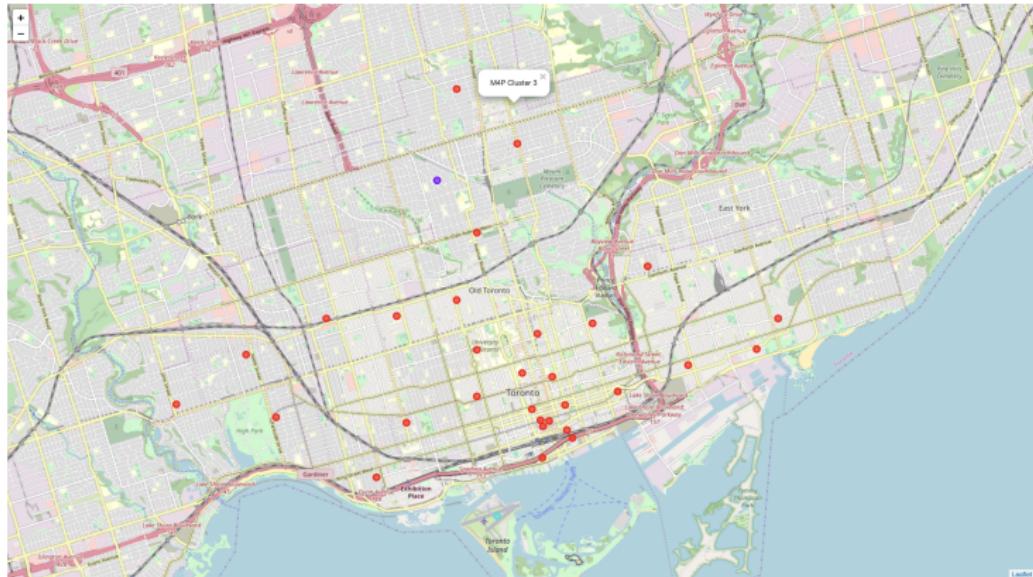


# Top 10 Most Common Food Categories



Coffee Shop is the most popular venue category in Toronto, followed by Café. The top two categories both serve coffee products mainly and are much more popular than the rest cuisines. Cuisines like Italian, Bakery, Pizza are also very popular, followed by Japanese and Sandwich.

# K-means clustering



- Cluster 0 (red): Coffee venues and restaurants
- Cluster 1 (purple): Cuisines with great diversity
- Cluster 2 (green): Best for brunch and tea time
- Cluster 3 (yellow): Fast-foods for quick meal or to go

# Discussion

- Coffee venues are most popular in most postal areas in Toronto.
- Most postal areas in Toronto are grouped into cluster 0, where coffee is most popular. These areas also have lots of other cuisines like Japanese and Italian.
- Cluster 1 - 3 are quite different from each other, where coffee is not very popular, and only one area is contained in each cluster.
- Cluster 1 has cuisines with great diversity.
- Cluster 0 & 1 are places to try cuisines from different countries and culture.
- Cluster 2 consists of venues best for brunch and tea time for people with free time to enjoy.
- Cluster 3 has lots of fast-foods venues, which is the place for busy people to grab some fast food to go.

# Conclusion

In this project, we clustered postal areas in Toronto based on popular cuisines in each area. We use different cuisines as feature and apply the k-means clustering algorithm to the data to group the areas into four clusters. Residence and travels can use the clusters to help them find the proper destination. For example, if someone is a lover of coffee, he should live in areas where can easily get a cup of coffee.