

# Introduction

This project analyzes neighborhoods in Toronto and tries to find the best place for families with small children. This cohort of people (families with children) has a specific list of requirements. The goal is to find the best fit.

## Business Problem

The family with children wants to move to Toronto and wants to buy the flat. By using data science methods we need to answer the question: “Which is the most suitable neighborhood in Toronto for families with small children.

## Data

The List of data:

- List of neighborhoods in Toronto, Canada
- Geodata of neighborhoods
- Venue category data

## Methodology

We get the list of neighborhoods in Toronto from the Wikipedia page: (“[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)”)

The data is then scrapped by utilizing BeautifulSoup.

The coordinates were taken from [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data). After it, we had all needed neighborhoods with their coordinates.

Foursquare is used to pull the list of venues near these neighborhoods.

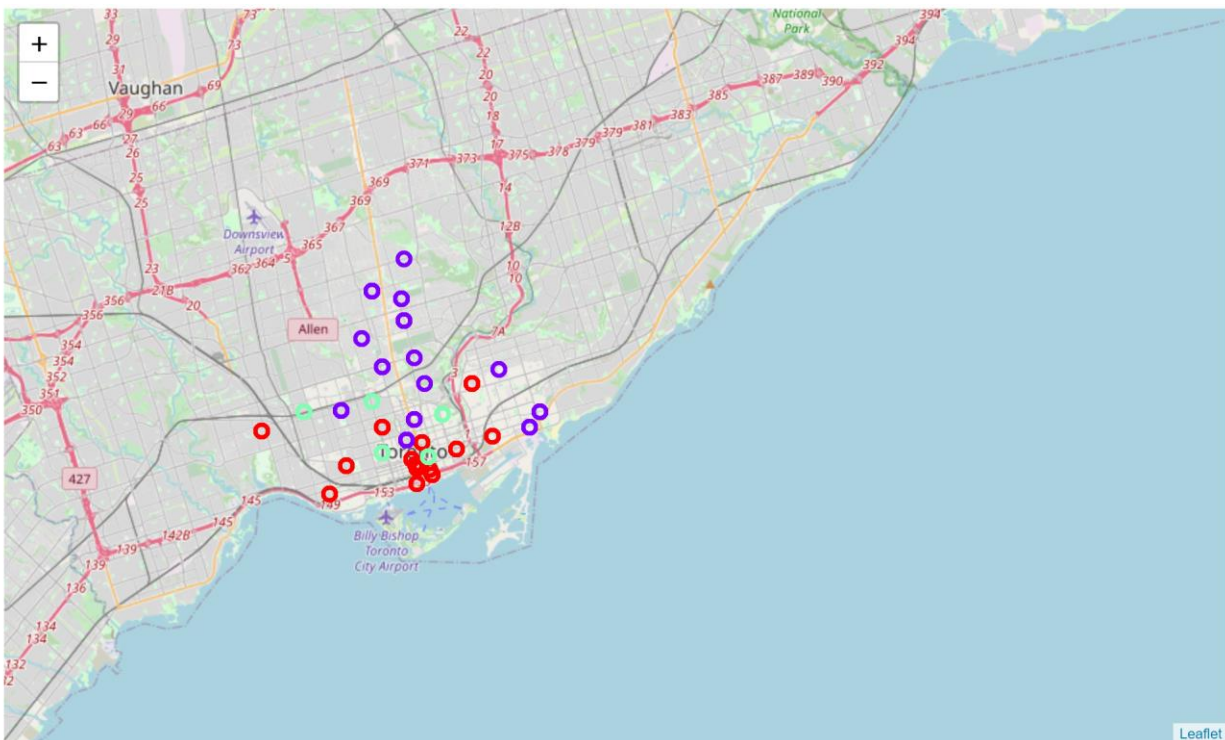
After gathering all these coordinates the Foursquare API is used to pull the list of the top 100 Venues within 500 meters radius.

The extracted data related to Bakery, Park, Grocery Store, Pharmacy, Playground, Baby Store, Supermarket, General Entertainment, Indoor Play Area. categories are considered as the basic requirements for families with children to live. Each

neighborhood is analyzed by grouping the rows by neighborhood and taking the occurrence of each venue category.

At last, we used the clustering method, k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster. The neighborhoods in Toronto were clustered into 3 clusters based on the occurrence of the venue categories. Based on the results (the concentration of clusters), this project is able to recommend the ideal location for families with children.

## Results



The results from k-means clustering show that we can categorize Toronto neighborhoods into 3 clusters based on the categories in each neighborhood:  
**Cluster 0 (red dots):** Cluster 0 has 15 Neighborhoods. This cluster has the biggest amount of neighborhoods, but mostly only Bakery and Parks are there.

**Cluster 1 (blue dots):** Cluster 1 has 14 Neighborhoods. Mostly only Parks are there.

**Cluster 2(green dots):** Neighborhoods from that cluster are the most suitable (Most of the requirements for families with children are met). Cluster 2 has 5 Neighborhoods.

## Recommendations

I would suggest choosing Neighborhoods from the **\*\*Cluster 2\*\***. Neighborhoods from that cluster are the most suitable.

For example **\*\*St. James Town, Cabbagetown\*\*** meets all the specified criteria (Bakery, Park, Grocery Store, Pharmacy, Playground, Baby Store, Supermarket, Fruit & Vegetable Store, Indoor Play Area) and will be comfortable for living with children.

If you pay attention to the map (cluster 0 has red color), then it can be noted that the districts from this cluster mostly are located in the city center.

## Conclusion

In this project, we identifying the real business problem, extracting and preparing the data, use machine learning by utilizing k-means clustering, and providing recommendations.