
Project Report: Battle of Neighborhoods

Alex J • 13/01/2010

Overview

- Grouping Toronto neighborhoods according to venues available
 - Use K-Means Cluster Analysis
 - Select the number of clusters
 - Produce descriptions of the Clusters
-

Business Problem

- Improve resident satisfaction through the reduction of travel time to places of interest.
 - 47% of Toronto households are renting, therefore are reasonably mobile.
 - Enabling the residents to choose locations that would optimise their access to their places of interest would not only improve their satisfaction, but would also reduce the pressure on transport infrastructure
-

Available Data

Administrative data sourced from Wikipedia.

- Postal Code, Borough, Neighborhood

Geo data

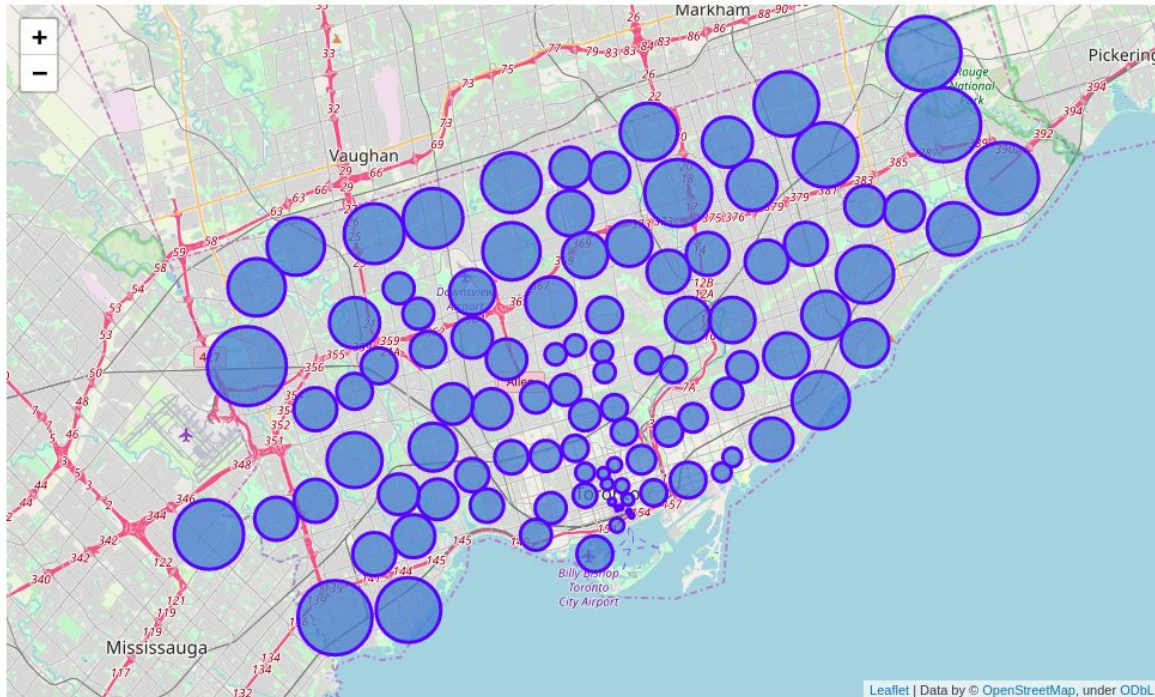
Geo data, available via the `**geopy**` package.

Foursquare data

We use Foursquare API to get information allowing us to assess the character of a neighborhood

Initial identification of neighborhoods

On the map of Toronto

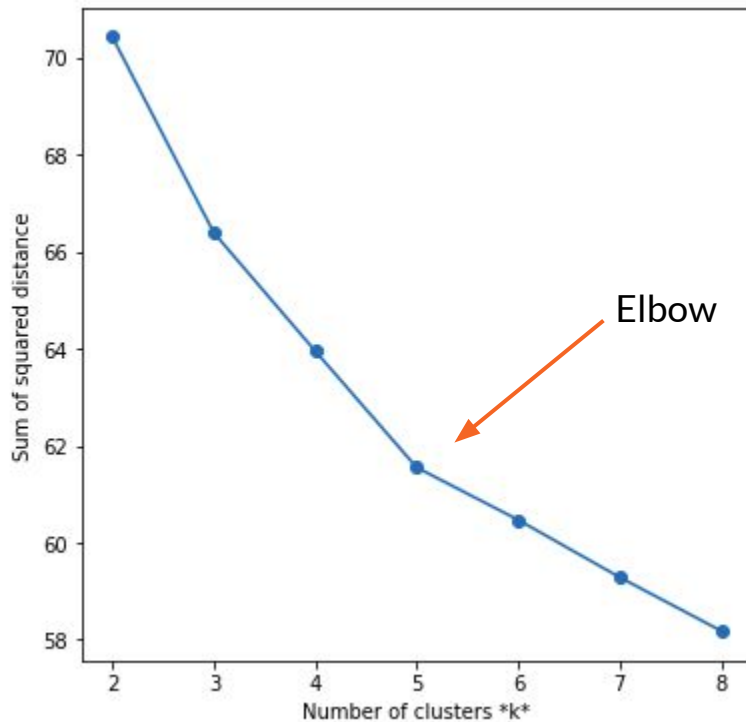


Clustering Analysis

Involved

- Getting necessary data from Foursquare
 - Producing a dataframe of venues per category
 - Normalising data to avoid huge distance between similar neighborhoods
 - Identifying optimal number of clusters
 - Producing descriptions of clusters as 4 most common categories of venues
-

Identifying optimal number of clusters



Map with Clusters

