Opening a Greek Grocery Store in Toronto.

Problem Statement

We’ve been hired by OPA ltd, a Greek Grocery store giant, to outsource a location in the Toronto Metropolitan Area, for starting up a Grocery Store Chain in Canada.

The Constraints to the problem were that it had to be for a busy enough area, able to support the venture, but not to be overcrowded, as overcrowded areas seem to be more competitive and infested with local crime, according to the OPA CEO’s personal prejudices. As such we have to model the problem with full endorsement of the client Constraints.

Data Sources

Wikipedia

Foursquare.

Dataset 1: Toronto postal codes, boroughs and neighborhoods

Data scraped with Python’s pandas and Wikipedia packages.

­Dataset 2: GPS neighborhood Data

Sourced from: <http://cocl.us/Geospatial_data>

Dataset 3: Toronto Venue Data from Foursquare API.

Data used to identify commonly visited venues in each of the Toronto neighborhoods. Identified most popular venues in each of the areas to satisfy the client’s requirement,

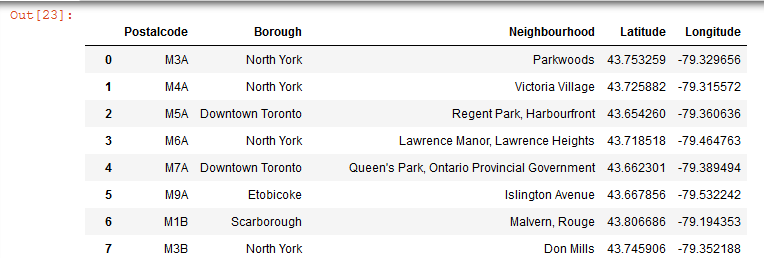
finally narrowing down the choice to just to proposed areas

Data Processing:

Cleansing Wikipedia data as per:

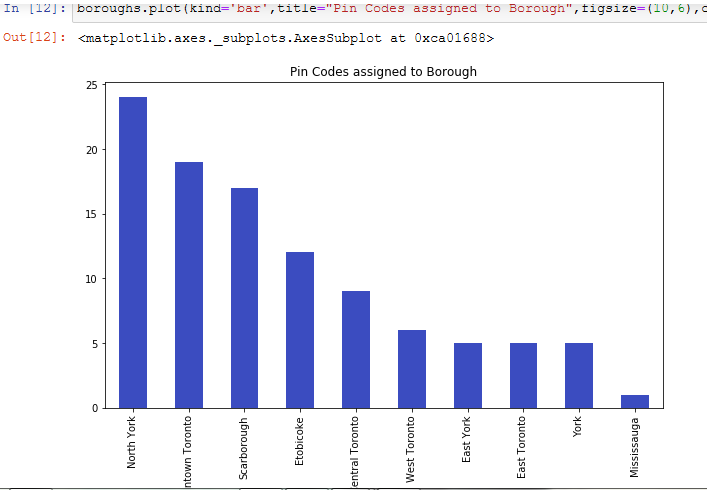
1. Dropping data lines where boroughs were not assigned
2. Combining different neighborhoods with the same postcode
3. Concatenating the GPS Data obtained from Dataset 2 to this dataframe

Sample:



Data Analysis

Analysis of pin codes associated with each Toronto borough.



Analysis of Toronto neighborhoods

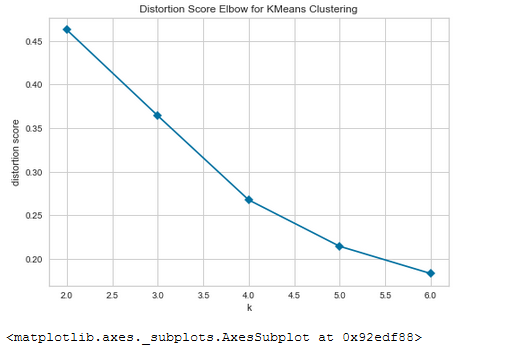
K-Means Clustering approach was used, utilizing the Foursquare API to extract the top 10 venues from each neighborhood.

Results:

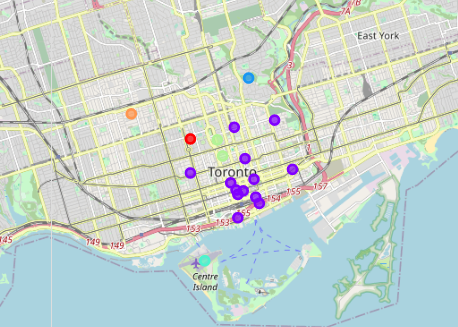


Clustering the neighborhoods

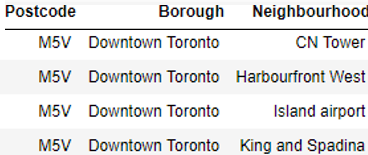
Deployed the elbow method to identify optimal clustering number



Followed by mapping:



Shortlisted neighborhoods in Cluster 1:



Discussion

Model can be improved by considering per capital income and crime index per Neighborhood