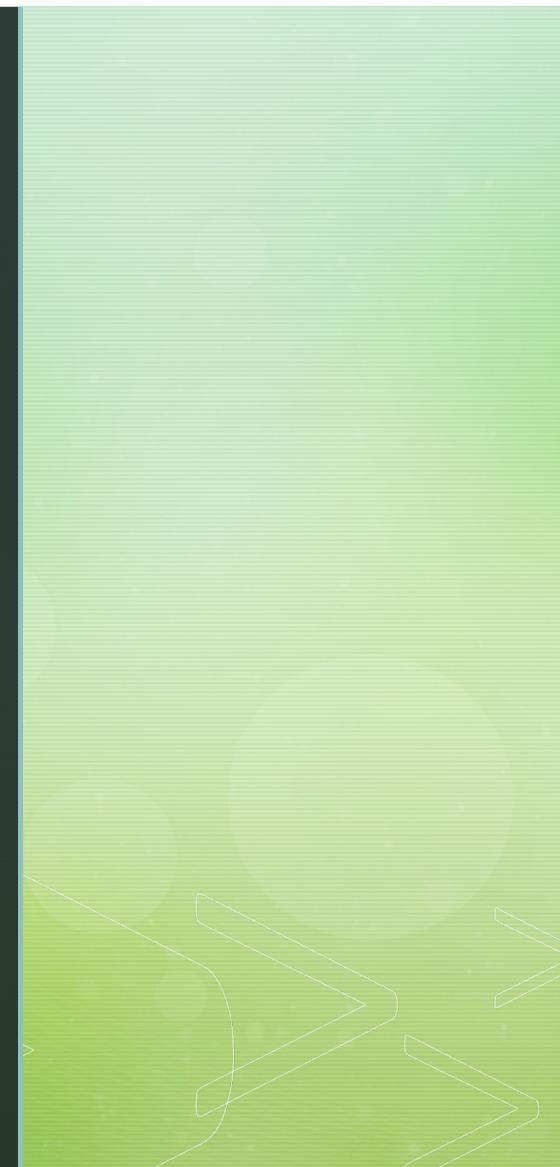


Understanding Neighborhoods

With Foursquare Venue Data

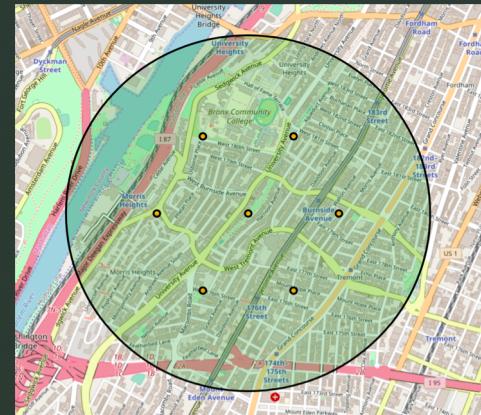


Data Enabled Relocation

- Provide a means to reduce the cost and effort needed to choose a suitable neighborhood in a new city
 - Find key targets for neighborhoods with similar structure to an existing neighborhood in another city
- Data to be used is the Foursquare venue database

Data Acquisition

- Data was collected by directing multiple queries in a 0.5 km radius circle around the geocoordinates of each neighborhood.
- Retrieved venues without coded categories were dropped.
- This left 140,227 venues in New York, and 71,040 venues in Toronto



A network of 7 query targets used to query the database

Data Engineering

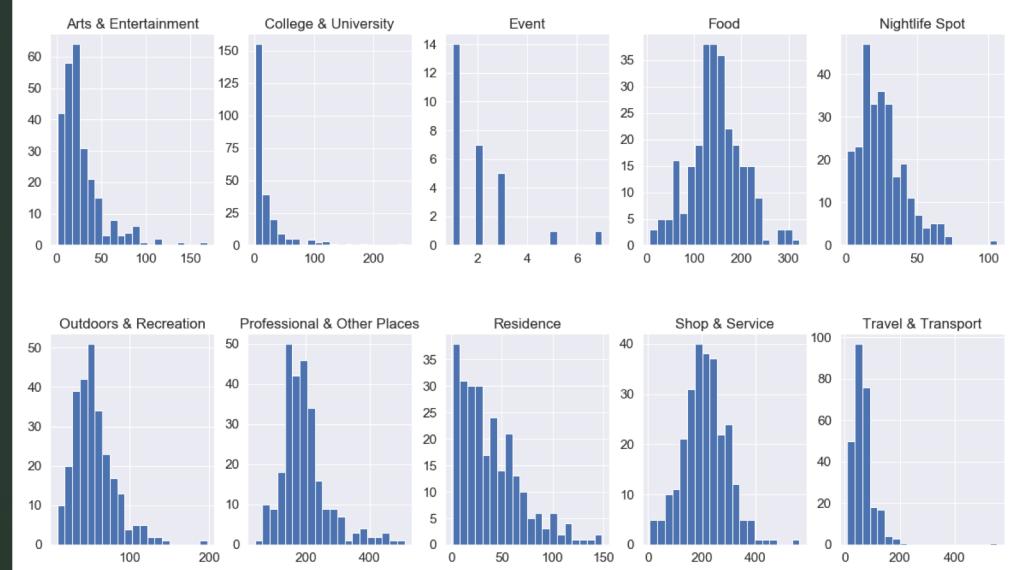
- A map of venue category to 10 category groups was developed from the hierarchical relationships published by FourSquare
- Venue categories within each category group were counted for each category group. These counts became the features used for cluster analysis.
- Aggregated category data was grouped and counted for each neighborhood, leaving a data set consisting of 270 neighborhoods, each with 10 features:

Arts & Entertainment	College and University	Event	Food	Nightlife Spot
Outdoors and Recreation	Professional and Other	Residence	Shop or Service	Travel and Transport

Data Exploration

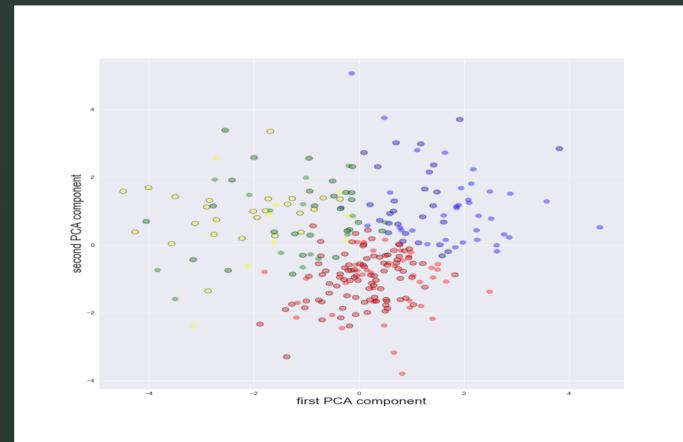
- Data for one feature, Events was found to be sparse and inconsistently used within the database.
- The Events Feature was dropped from analysis

Histograms of Feature Values (venue counts)

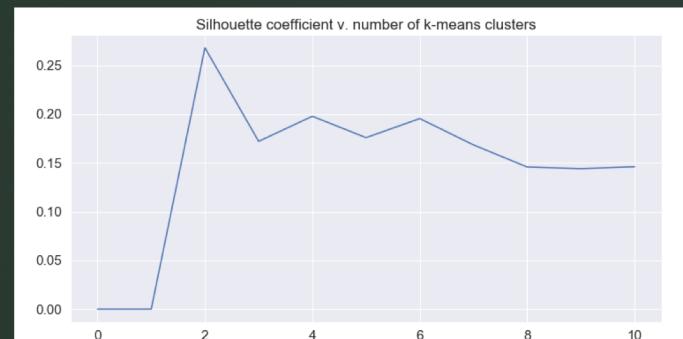


Data Analysis

- Clustering of Neighborhoods provided some value, removing 65% + of neighborhoods as candidates.
- Results still left Too many neighborhoods to provide a concise relocation search.
- In Addition the resulting silhouette coefficients, all less than 0.20 for $3 \leq k \leq 5$, suggested that the resulting clusters were based on insufficient structure found in the underlying dataset



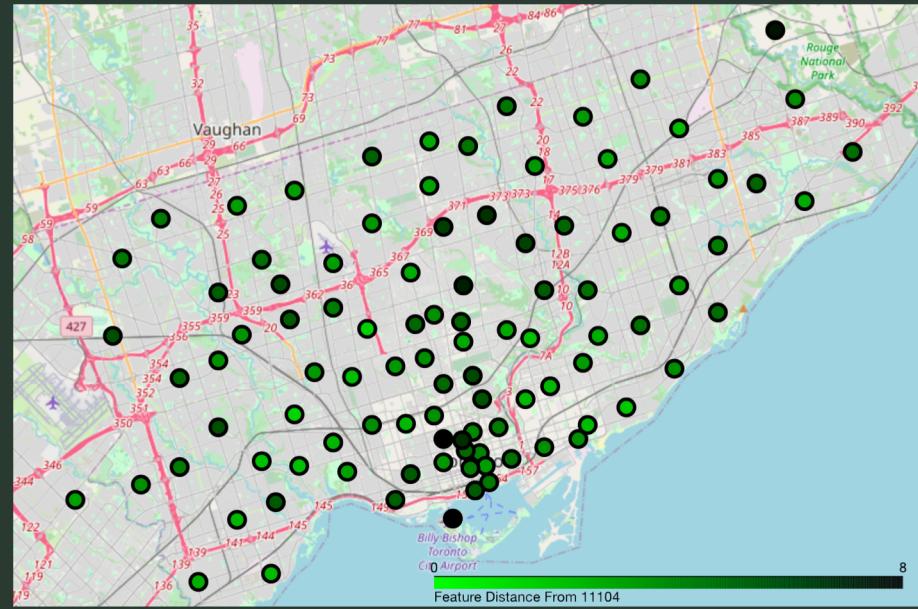
Clusters: k=4 (2 Component PCA projection)



Average silhouette coefficients, k=2 to k=10

Alternate Approach

- Absolute Euclidean feature distance was used to compare neighborhoods
- The feature distance from the home neighborhood was calculated for all Toronto Neighborhoods
- The results were plotted onto a map of Toronto, with best candidate neighborhoods visualized in bright green.



Euclidean Distance for Toronto Neighborhoods, based on features scaled to $\mu=0$, $\sigma = 1$. Neighborhoods that closely match the home neighborhood show as brighter green