# Capstone – Recommending Relocation Neighborhood

BY ILYA TURIAN

# Introduction

An Imaginary Software Company has offices in 3 cities:

- New-York
- London
- Toronto

Due to business requirements, employees are often required to relocate between those cities.

# Introduction

- The company has a great interest in keeping it’s employees happy and therefore the HR department provides assistance in the relocation process.
- The most common questions asked are related to finding housing in the new town. And specifically :

**“In which neighborhood in the new area should I look for apartment?”**

# “In which neighborhood in the new city should I look for apartment?”

- To answer the above question we will resort to Data Science and machine learning.
- We are basing our work on the following assumptions:
  - The employee would like to live in an area similar to the area he currently live in.
  - We can get a good characteristic of an area by scrapping Foursquare data and finding the most common venues in that area. This data should represent the general atmosphere and type of the neighborhood.

# “In which neighborhood in the new city should I look for apartment?”

The general plan is to gather Foursquare data for each area. Normalize it and then use a clustering algorithm to cluster the neighborhoods from different cities.

Using this model, we can see in which cluster is the current residence of the employee and suggest similar areas in a different city.

# Data Sources

To differentiate the neighborhoods we need to get coordinates for each neighborhood in each city.

- **Toronto:**

  Scrapping **-** Wikipedia - Toronto Postal Codes

- **New-York:**

  Geospatial_Coordinates.csv provided in one of the labs.

- **London:**

  **https://tools.wmflabs.org/kmlexport?article=Category%3AAreas+of+London**

# Data Sources

The Foursquare data used to characterize each neighborhood will be pulled through the Foursquare developer API, documented at: https://developer.foursquare.com/docs

# Methodology – Neighborhood Data

A table containing all the Neighborhoods was created by scrapping the coordinates source and merging.

The list was narrowed down by focusing on central areas for each cities since Foursquare data is really lacking for suburbs.

| | Neighbourhood | Latitude | Longitude | City |
|---|---|---|---|---|
| 1 | Acton, London | 51.513519 | -0.270661 | London |
| 2 | Acton Green, London | 51.510515 | -0.262668 | London |
| 3 | Acton Vale, London | 51.511000 | -0.258000 | London |
| 4 | Addington, London | 51.358300 | -0.030500 | London |
| 6 | Adelphi, London | 51.509167 | -0.122500 | London |

# Methodology – Foursquare Scrapping

- For each neighborhood center we got all the venues in a 500 yard radius from it center.
- We calculated the ratio of each venue type in it's neighborhood:

| | Neighbourhood | Accessories Store | Adult Boutique | Afghan Restaurant | African Restaurant | American Restaurant | Animal Shelter | Antique Shop | Arcade | Arepa Restaurant | ... | Volleyball Court | Watch Shop | Waterfront | Weight Loss Center |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Battery Park City | 0.0 | 0.0 | 0.0 | 0.00000 | 0.010000 | 0.0 | 0.00 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |
| 1 | Carnegie Hill | 0.0 | 0.0 | 0.0 | 0.00000 | 0.010000 | 0.0 | 0.00 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |
| 2 | Central Harlem | 0.0 | 0.0 | 0.0 | 0.06383 | 0.042553 | 0.0 | 0.00 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |
| 3 | Chelsea | 0.0 | 0.0 | 0.0 | 0.00000 | 0.030000 | 0.0 | 0.01 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |
| 4 | Chinatown | 0.0 | 0.0 | 0.0 | 0.00000 | 0.040000 | 0.0 | 0.00 | 0.0 | 0.0 | ... | 0.0 | 0.0 | 0.0 | 0.0 |

- Finally, we merged the 3 data frames into a single data frame containing neighborhoods from the 3 cities.

# Methodology – Dealing with NaN's

- After the merge, we encountered quite a few NaN features in our data frame.
- The reason for this was some venue types that are common in some cities, but non existent in other.
- American Restaurant, for example, exist in all 3 cities, but Antique Shop is not.

| Venue Type | Number of Nan's |
| --- | --- |
| Accessories Store | 38 |
| Adult Boutique | 138 |
| Afghan Restaurant | 0 |
| African Restaurant | 38 |
| Airport | 178 |
| Airport Food Court | 178 |
| Airport Gate | 178 |
| Airport Lounge | 178 |
| Airport Service | 178 |
| Airport Terminal | 178 |
| American Restaurant | 0 |
| Animal Shelter | 176 |
| Antique Shop | 138 |

# Methodology – Dealing with NaN’s

- Two approaches were considered:
  A. Dropping the columns (Venue Types) that don’t exist in the 3 cities.
  B. Replacing the Nan’s with 0.
- The second approach was chosen due to 2 factors:
  A. Not having a certain venue type, is also a feature that should be considered.
  B. The accuracy metrics were better using the second solution.

# Methodology – Method

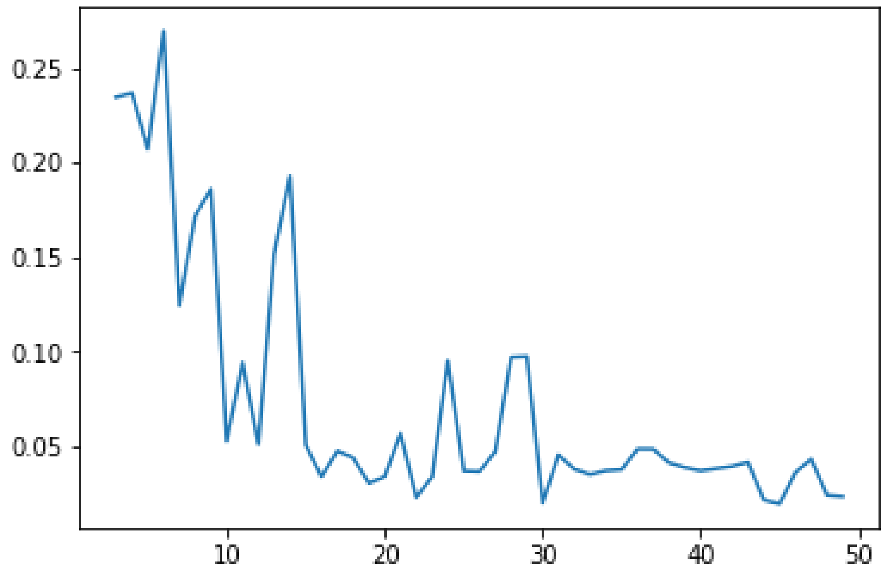
- To achieve our business goal, we decided to use K-nearest neighbor classifier.
- This is an unsupervised training method that groups members by their feature similarity, in our case, ratio of certain venues.

# Methodology – choosing k value

- Before our training we must decide on the k value – the amount of clusters we want to be created.
- To do this we used the "Elbow Method". We trained the model using different values of k and looked for the best accuracy.
- We received the following graph:
- The absolute best was at k=7, however

  we decided it's quite low and won't

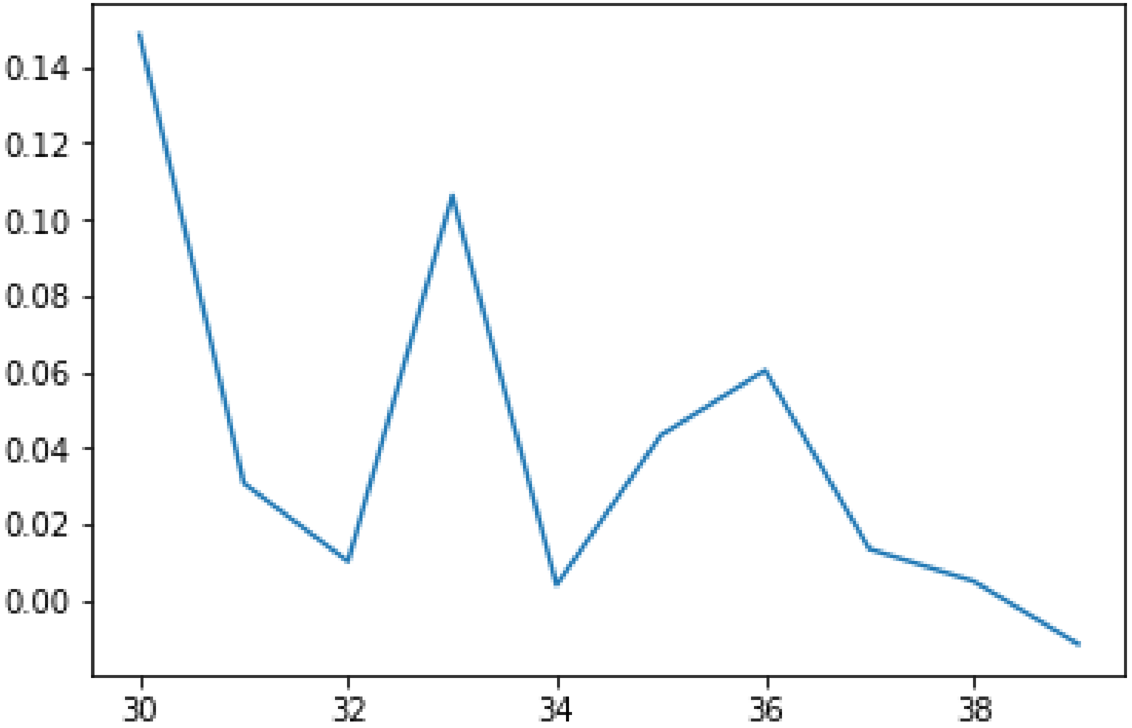
  create good enough differentiation.



# Methodology – choosing k value

- A local maximum was observed in the ~32 area, so we decided to zoom the graph on that area.
- After some Testing 29 was chosen as

  since it yielded the best clusters.



# Methodology - Validation

- To validate our results, we used common sense by checking the 10 most common venues in each neighborhood and see if they make sense in "sister neighborhoods" concept we planned:.

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Battery Park City | Park | Coffee Shop | Hotel | Memorial Site | Wine Shop | Italian Restaurant | Clothing Store | Gym | Plaza | Burger Joint |
| **1** | Carnegie Hill | Pizza Place | Coffee Shop | Café | Wine Shop | Cosmetics Shop | Bar | Japanese Restaurant | Spa | French Restaurant | Grocery Store |
| **2** | Central Harlem | Cosmetics Shop | African Restaurant | Art Gallery | French Restaurant | Seafood Restaurant | Gym / Fitness Center | Public Art | Chinese Restaurant | American Restaurant | Event Space |

# Results - Clusters

- After the training, we checked to see neighborhood distribution in each cluster.
- Unfourtentley we have created only 6 useful clusters – (0,1,4,6,7,26).
- The other clusters contained neighborhoods that are not similar enough to be clustered.
- Let’s check and characterize each of “interesting” cluster.

| Cluster | City | Neighbourhood |
|---|---|---|
| 0 | London | 17 |
| | NY | 30 |
| | Toronto | 5 |
| 1 | London | 4 |
| | Toronto | 1 |
| 2 | London | 1 |
| 3 | London | 1 |
| 4 | London | 5 |
| | Toronto | 1 |
| 5 | London | 1 |
| 6 | London | 10 |
| | NY | 10 |
| | Toronto | 6 |
| 7 | London | 28 |
| | Toronto | 19 |
| 8 | London | 1 |
| 9 | Toronto | 1 |
| 10 | London | 8 |
| 11 | Toronto | 1 |
| 12 | London | 1 |
| 13 | London | 1 |
| 14 | London | 1 |
| 15 | Toronto | 1 |
| 16 | London | 5 |
| 17 | London | 1 |
| 18 | London | 4 |
| 19 | London | 13 |
| 20 | London | 12 |
| 21 | Toronto | 1 |
| 22 | London | 2 |
| 23 | London | 1 |
| 24 | London | 1 |
| 25 | London | 2 |
| 26 | London | 16 |
| | Toronto | 2 |
| 27 | London | 1 |
| 28 | London | 1 |

# Results – Cluster 0 – Asian Style

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | Church and Wellesley | Japanese Restaurant | Coffee Shop | Sushi Restaurant | Restaurant | Gay Bar | Men's Store | Gym | Pub | Mediterranean Restaurant | Bubble Tea Shop | T |
| 16 | Kensington Market, Chinatown, Grange Park | Café | Vegetarian / Vegan Restaurant | Bakery | Mexican Restaurant | Coffee Shop | Dumpling Restaurant | Bar | Vietnamese Restaurant | Chinese Restaurant | Cocktail Bar | T |
| 17 | King, Adelaide, Richmond | Coffee Shop | Café | Steakhouse | Bar | American Restaurant | Bakery | Restaurant | Hotel | Burger Joint | Cosmetics Shop | T |
| 20 | Railway Lands, King and Spadina, CN Tower, Isl... | Airport Terminal | Airport Lounge | Airport Service | Plane | Sculpture Garden | Boat or Ferry | Boutique | Harbor / Marina | Bar | Airport Gate | T |
| 35 | Trinity, Little Portugal | Bar | Asian Restaurant | Men's Store | Coffee Shop | Pizza Place | New American Restaurant | Restaurant | Vietnamese Restaurant | Café | Cocktail Bar | T |
| 39 | Carnegie Hill | Coffee Shop | Pizza Place | Café | Yoga Studio | Bookstore | Wine Shop | Cosmetics Shop | French Restaurant | Bar | Japanese Restaurant | |
| 40 | Central Harlem | African Restaurant | Public Art | Art Gallery | Seafood Restaurant | Chinese Restaurant | Gym / Fitness Center | French Restaurant | American Restaurant | Cosmetics Shop | Liquor Store | |
| 41 | Chelsea | Coffee Shop | Ice Cream Shop | Italian Restaurant | Bakery | Nightclub | Theater | Seafood Restaurant | American Restaurant | Hotel | Art Gallery | |
| 42 | Chinatown | Chinese Restaurant | American Restaurant | Cocktail Bar | Spa | Dumpling Restaurant | Vietnamese Restaurant | Bubble Tea Shop | Optical Shop | Salon / Barbershop | Ice Cream Shop | |

# Results – Cluster 1 – Quite Living

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | Christie | Café | Grocery Store | Park | Convenience Store | Coffee Shop | Baby Store | Nightclub | Diner | Italian Restaurant | Restaurant | Toronto |
| 91 | Bellingham, London | Grocery Store | Gym | Park | Train Station | Falafel Restaurant | Farm | Farmers Market | Fast Food Restaurant | Field | Film Studio | London |
| 116 | East End of London | Park | Grocery Store | Bakery | Fried Chicken Joint | Thrift / Vintage Store | Farm | Indian Restaurant | Flower Shop | Pub | Chinese Restaurant | London |
| 163 | Northfields, London | Grocery Store | Italian Restaurant | Park | Café | Pub | Chinese Restaurant | Cheese Shop | French Restaurant | Deli / Bodega | Kebab Restaurant | London |
| 186 | St Ann's, London | Park | Café | Grocery Store | Hostel | Fish & Chips Shop | Falafel Restaurant | Farm | Farmers Market | Fast Food Restaurant | Field | London |

# Results – Cluster 4 – Country Side

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 18 | Lawrence Park | Park | Swim School | Bus Line | Yoga Studio | Doner Restaurant | Fish & Chips Shop | Filipino Restaurant | Fast Food Restaurant | Farmers Market | Falafel Restaurant | Toronto |
| 85 | Ashburton, London | Tram Station | Park | Chinese Restaurant | African Restaurant | Yoshoku Restaurant | Fish & Chips Shop | Farmers Market | Fast Food Restaurant | Field | Film Studio | London |
| 133 | Hampton, London | Park | Fried Chicken Joint | Coffee Shop | Yoshoku Restaurant | Farm | Farmers Market | Fast Food Restaurant | Field | Film Studio | Fish & Chips Shop | London |
| 153 | Manor Park, London | Gas Station | Restaurant | Fried Chicken Joint | Turkish Restaurant | Park | Film Studio | Falafel Restaurant | Farm | Farmers Market | Fast Food Restaurant | London |
| 157 | Mitcham, London | Construction & Landscaping | Bus Stop | Bookstore | Park | Yoshoku Restaurant | Farmers Market | Fast Food Restaurant | Field | Film Studio | Fish & Chips Shop | London |
| 165 | Northwood, London | Park | Soccer Field | Gym / Fitness Center | Golf Course | Farm | Farmers Market | Fast Food Restaurant | Field | Film Studio | Yoshoku Restaurant | London |

# Results – Cluster 6 – Upscale Housing

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Business Reply Mail Processing Centre 969 Eastern | Yoga Studio | Auto Workshop | Comic Shop | Pizza Place | Recording Studio | Restaurant | Butcher | Burrito Place | Skate Park | Brewery | Toronto |
| 6 | Davisville North | Playground | Hotel | Clothing Store | Food & Drink Shop | Grocery Store | Park | Gym | Breakfast Spot | Sandwich Place | Falafel Restaurant | Toronto |
| 8 | Dufferin, Dovercourt Village | Supermarket | Pharmacy | Bakery | Brazilian Restaurant | Bank | Brewery | Music Venue | Discount Store | Coffee Shop | Café | Toronto |
| 15 | India Bazaar, The Beaches West | Park | Gym | Italian Restaurant | Pizza Place | Pub | Movie Theater | Sandwich Place | Burrito Place | Burger Joint | Brewery | Toronto |
| 19 | North Toronto West | Coffee Shop | Yoga Studio | Bagel Shop | Park | Clothing Store | Dessert Shop | Chinese Restaurant | Rental Car Location | Diner | Salon / Barbershop | Toronto |
| 34 | The Junction South, High Park | Bar | Mexican Restaurant | Café | Fast Food Restaurant | Fried Chicken Joint | Bakery | Italian Restaurant | Gastropub | Music Venue | Arts & Crafts Store | Toronto |
| 38 | Battery Park City | Park | Coffee Shop | Hotel | Memorial Site | Wine Shop | Italian Restaurant | Clothing Store | Gym | Plaza | Men's Store | NY |
| 45 | East Harlem | Mexican Restaurant | Deli / Bodega | Bakery | Latin American Restaurant | Thai Restaurant | Convenience Store | Café | Gas Station | Taco Place | Steakhouse | NY |

# Results – Cluster 7 – Hip places

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Berczy Park | Coffee Shop | Cocktail Bar | Steakhouse | Bakery | Café | Cheese Shop | Seafood Restaurant | Beer Bar | Italian Restaurant | Farmers Market | Toronto |
| **2** | Central Bay Street | Coffee Shop | Café | Italian Restaurant | Middle Eastern Restaurant | Sandwich Place | Burger Joint | Japanese Restaurant | Ice Cream Shop | Spa | Bar | Toronto |
| **5** | Davisville | Pizza Place | Dessert Shop | Sandwich Place | Italian Restaurant | Café | Thai Restaurant | Sushi Restaurant | Coffee Shop | Restaurant | Deli / Bodega | Toronto |
| **7** | Design Exchange, Toronto Dominion Centre | Coffee Shop | Café | Hotel | Restaurant | Gastropub | Deli / Bodega | Italian Restaurant | Bakery | American Restaurant | Pizza Place | Toronto |
| **9** | Exhibition Place, Brockton, Parkdale Village | Breakfast Spot | Café | Coffee Shop | Yoga Studio | Intersection | Performing Arts Venue | Caribbean Restaurant | Stadium | Restaurant | Bar | Toronto |
| **10** | First Canadian Place, Underground city | Coffee Shop | Café | Hotel | Steakhouse | Restaurant | Bar | Deli / Bodega | Seafood Restaurant | Gastropub | American Restaurant | Toronto |
| **12** | Garden District, Ryerson | Coffee Shop | Clothing Store | Cosmetics Shop | Café | Middle Eastern Restaurant | Tea Room | Italian Restaurant | Diner | Pizza Place | Bubble Tea Shop | Toronto |

# Results – Cluster 26 – Business Areas

| | Neighbourhood | 1th Most Common Venue | 2th Most Common Venue | 3th Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue | City |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **25** | South Hill, Rathnelly, Forest Hill SE, Deer Pa... | Pub | Coffee Shop | Liquor Store | Light Rail Station | Sushi Restaurant | Supermarket | Sports Bar | Fried Chicken Joint | American Restaurant | Vietnamese Restaurant | Toronto |
| **32** | The Beaches | Health Food Store | Pub | Trail | Neighborhood | Other Great Outdoors | Falafel Restaurant | Event Space | Farmers Market | Fast Food Restaurant | Filipino Restaurant | Toronto |
| **78** | Acton Green, London | Pub | Grocery Store | Park | Wine Shop | Creperie | Bakery | Mini Golf | Convenience Store | Gym / Fitness Center | Train Station | London |
| **79** | Acton Vale, London | Gym / Fitness Center | Pub | Park | Chinese Restaurant | Bakery | Train Station | Mini Golf | Yoshoku Restaurant | Farmers Market | Fast Food Restaurant | London |
| **102** | Castle Green, London | Skate Park | Go Kart Track | Pub | Rugby Pitch | Field | Exhibit | Falafel Restaurant | Farm | Farmers Market | Fast Food Restaurant | London |
| **125** | Forest Hill, London | Pub | Coffee Shop | Café | Grocery Store | Gym / Fitness Center | Train Station | Bookstore | Indian Restaurant | Thai Restaurant | Gastropub | London |
| **128** | Fulwell, London | Pizza Place | Bus Station | Golf Course | Pub | Diner | Seafood Restaurant | Convenience Store | Fast Food Restaurant | Chinese Restaurant | Garden Center | London |
| **134** | Harlington, London | Restaurant | Indian Restaurant | Pub | Bus Stop | Gym / Fitness Center | Rental Car Location | Café | Hotel | Pool | Grocery Store | London |

# Conclusions

- We had a really partial success.
- The clustering didn't manage to classify all the neighborhoods and a lot of them were stuck in a single cluster.
- The Clusters that did contain several neighborhoods were relatively cohesive and can be used to partly answer the initial question.
- More features are needed to create a better clustering.