# Finding similar neighbourhood across Cities based on Venues around

**The Battle of Neighbourhoods**

**30 April 2020**

Shreeshyla H S

Software Engineer

# 1. Introduction:

## 1.1 Background:

Often people have to relocate to different cities due to Job change, there is always a confusion on which neighborhood to shift to in a particular city. Varieties of questions arise, Should I find the neighborhood which is closer to the new workplace? Should I find the neighborhood which is similar to my current neighborhood? Should I explore some unique neighborhoods around? Does the cost of living matter there? Or Does the crime rate matter compared to my current neighborhood? and so many different questions arise as one starts exploring the place to relocate.

## 1.2 Business Problem:

As we can get the neighbourhood data of every city, the venues around the neighbourhood, the distance between neighbourhood and venues along with rating of venues, this project aims to predict the similar neighbourhood based on the current neighbourhood and also the nearest to the new desired place of work.

## 1.3 Usage:

Almost all the working professionals who desire to relocate to the new city for better opportunity to work, need a tool which suggests to them the neighbourhood based on their interest in living.

As an instance, a Person named "Adam" living in "St. James Town, Downtown Toronto, Canada" has earned a Job in "Midtown south, Manhattan, New York". The proposed model will suggest the neighborhood of Manhattan which is similar to Adam's current neighborhood and also the nearest to his new workplace.

# 2. Data Acquisition:

### 2.1 City Data:

Canada:

Data Source: Canada Borough Data from Wiki

Following attribute is fetched from the data source:

- postal code
- Neighbourhood
- Borough

Latitude and Longitude of the location can be fetched from the postal code.

For this project, "St. James Town, Downtown Toronto, Canada" is considered as the source city the user is relocating from.

| Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|
| Downtown Toronto | St. James Town | 43.651494 | -79.375418 |

New York:

Data Source: Curated list of NewYork Neighbourhoods

Manhattan, Newyork is considered as our destination city for this project.

Below you can check the sample of data,

```
      Borough        Neighborhood   Latitude  Longitude
0   Manhattan         Marble Hill  40.876551 -73.910660
1   Manhattan           Chinatown  40.715618 -73.994279
2   Manhattan  Washington Heights  40.851903 -73.936900
3   Manhattan             Inwood  40.867684 -73.921210
4   Manhattan    Hamilton Heights  40.823604 -73.949688

 Manhattan has 40 neighborhoods
```

### 2.2 Foursquare API

The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag.

Foursquare API gives us the following data:

- List of Venues across Neighborhood
- Venue ratings
- Venue Check-ins

We propose a method to solve the problem by finding similar venues across the source and destination neighborhoods. Checking against the time of Check-ins & the location entropy across the nearby similar venues and also considering the geographical space/distance from the desired location of work, we can suggest the set of Neighborhood the user can relocate to.

You can check below the output **Figure 2.1** of FourSquare API where it gives out the number of venues in each neighbourhood and number of unique categories in the data set **Figure 2.2**.

```
Below are the Venue information for the respective Neighborhoods:
Marble Hill : 25 Venues
Chinatown : 100 Venues
Washington Heights : 90 Venues
Inwood : 57 Venues
Hamilton Heights : 59 Venues
Manhattanville : 46 Venues
Central Harlem : 42 Venues
East Harlem : 42 Venues
Upper East Side : 90 Venues
Yorkville : 100 Venues
Lenox Hill : 100 Venues
Roosevelt Island : 25 Venues
Upper West Side : 68 Venues
Lincoln Square : 95 Venues
Clinton : 100 Venues
Midtown : 100 Venues
Murray Hill : 87 Venues
Chelsea : 100 Venues
Greenwich Village : 100 Venues
East Village : 100 Venues
Lower East Side : 47 Venues
Tribeca : 74 Venues
Little Italy : 100 Venues
Soho : 74 Venues
West Village : 100 Venues
Manhattan Valley : 40 Venues
Morningside Heights : 39 Venues
Gramercy : 72 Venues
Battery Park City : 60 Venues
Financial District : 100 Venues
Carnegie Hill : 86 Venues
Noho : 100 Venues
Civic Center : 87 Venues
Midtown South : 87 Venues
Sutton Place : 93 Venues
Turtle Bay : 100 Venues
Tudor City : 73 Venues
Stuyvesant Town : 15 Venues
Flatiron : 100 Venues
Hudson Yards : 55 Venues
```

**Figure 2.1**

```
There are 321 uniques categories.


0          Pizza Place
1          Yoga Studio
2                Diner
3           Donut Shop
4          Coffee Shop
              ...
316    Baseball Field
317       Sports Club
318      Tech Startup
319      Camera Store
320           Stables
```

**Figure 2.2**

# 3. Exploratory Data Analysis:

## 3.1 Venues across source Neighbourhood

As we have selected source neighbourhood as St.James Town, Downtown Toronto, Canada, we fetch the venues around the neighborhood and we can visualize it in the map using folium library as seen below **(Figure 3.1)**,

```
Below are the Venue information for the respective Neighborhoods:
St. James Town : 80 Venues
There are 54 uniques categories.
```

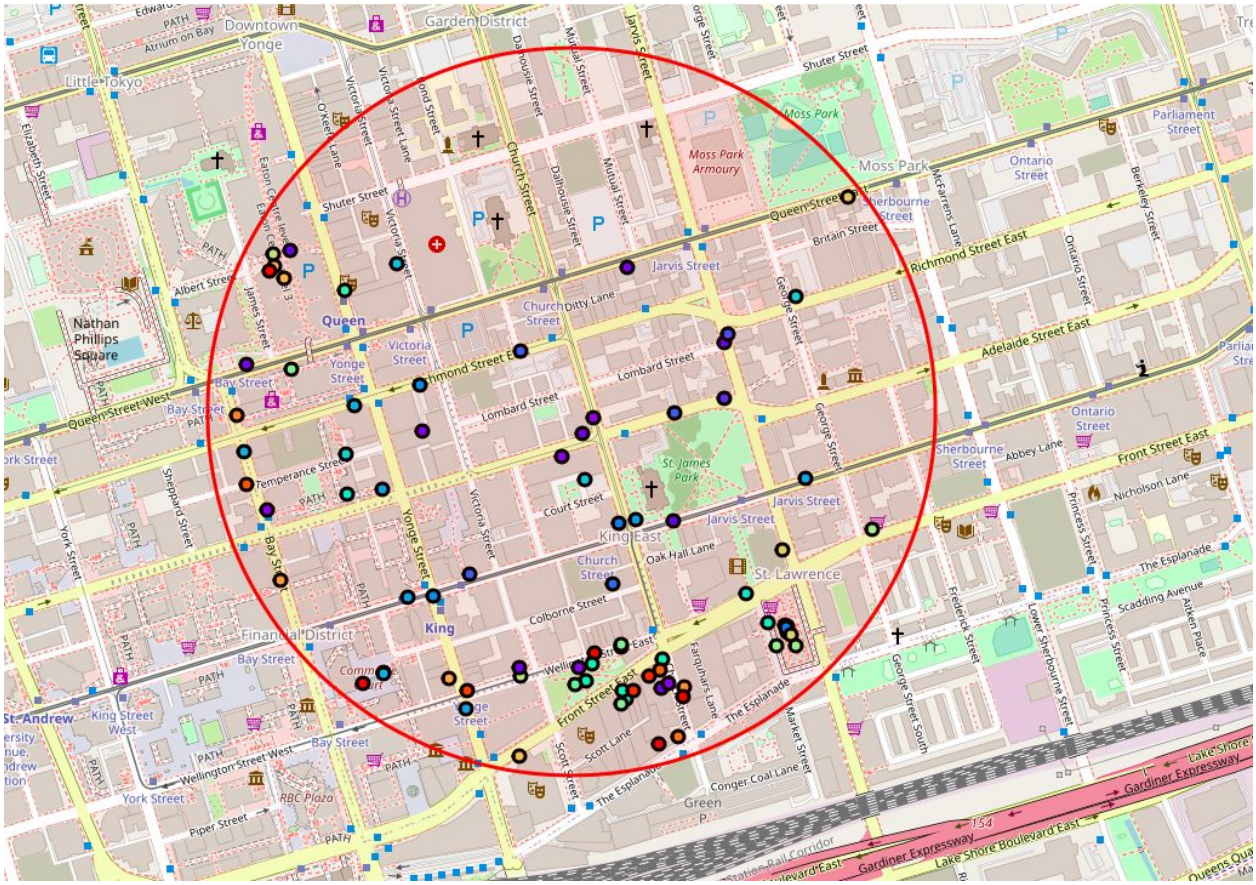


**Figure 3.1**

Below is the bar chart (**Figure 3.2**) for the Venues across the Source Neighbourhood for better visualisation on how many Venues exist.

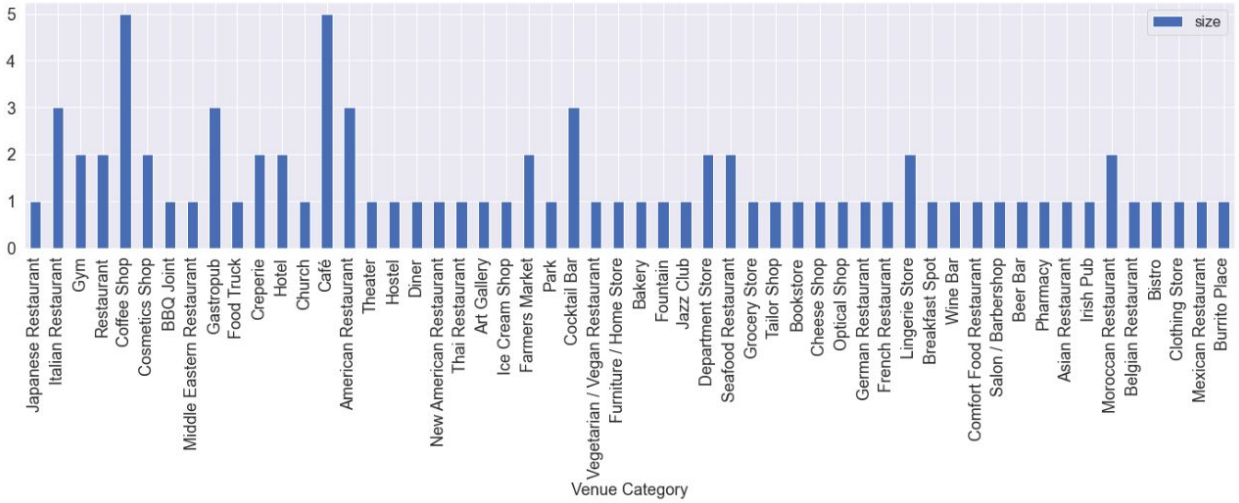


**Figure 3.2**

As you can see, the Neighbourhood is ruling with different genres of restaurant and other food services, Cafes, Gym, Church and other Miscellaneous venues.

Below Pie Chart depicting the 20 most common venues and is self-explanatory,

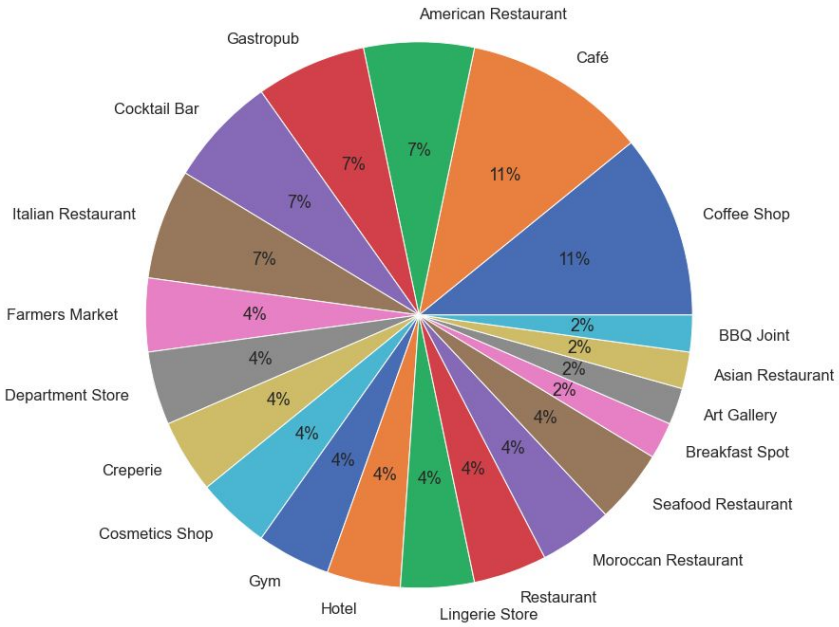


**Figure 3.3**

## 3.2 Venues across destination Borough

Analysing 40 Neighbourhoods around Manhattan Borough using Folium, **Figure 3.4** depicts the Neighbourhoods.

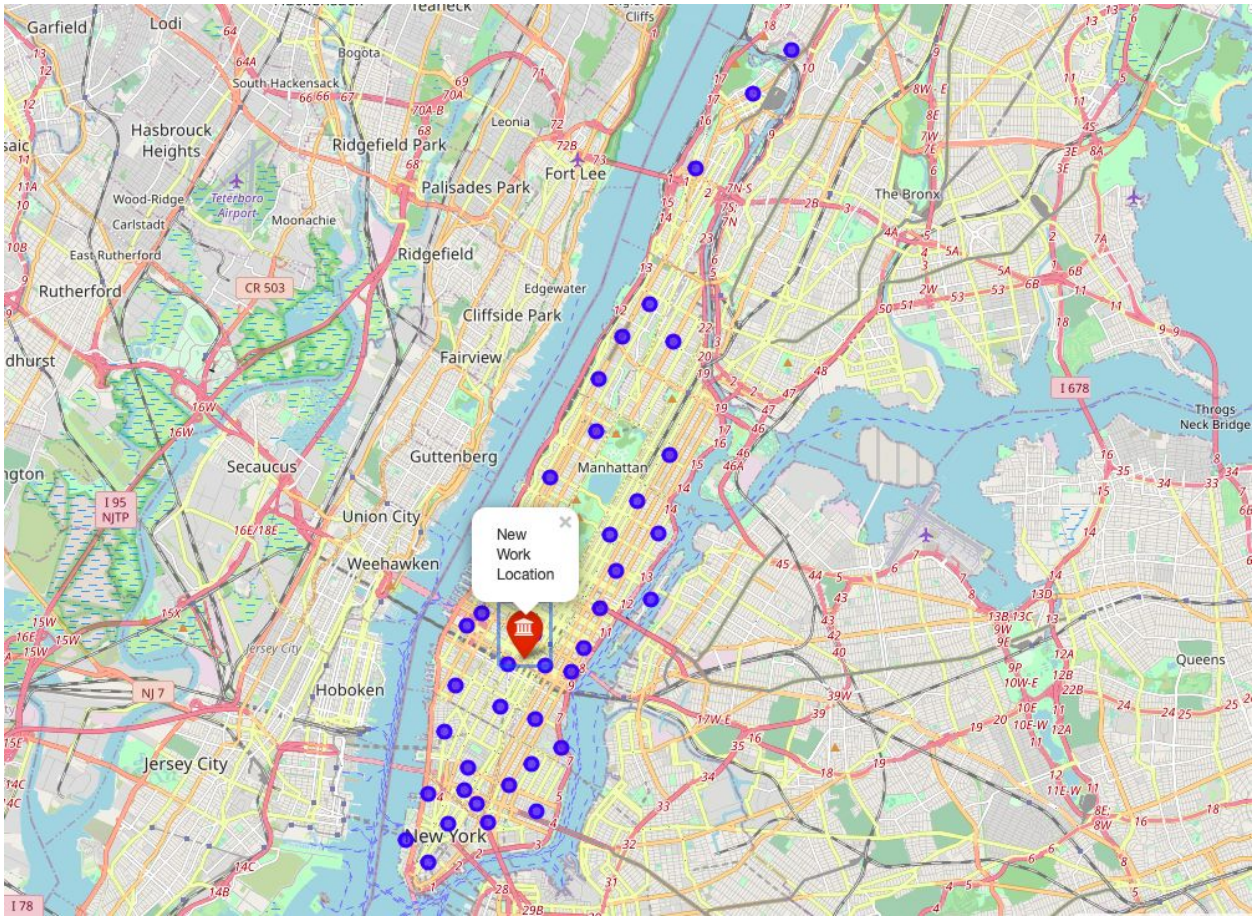


**Figure 3.4**

Below shown table (**Figure 3.5**) depicts 20 most common Venues around the above 40 Neighbourhoods,

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd 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 | ... | 11th Most Common Venue | 12th Most Common Venue | 13th Most Common Venue | 14th Most Common Venue | 15th Most Common Venue | 16th Most Common Venue | 17th Most Common Venue | 18th Most Common Venue | 19th Most Common Venue | 20th Most Common Venue |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Battery Park City | Park | Hotel | Gym | Memorial Site | Burger Joint | Wine Shop | Department Store | Italian Restaurant | Plaza | ... | Gourmet Shop | Food Court | Shopping Mall | Coffee Shop | Building | Lingerie Store | Electronics Store | Chinese Restaurant | Sandwich Place | Bistro |
| 1 | Carnegie Hill | Coffee Shop | Café | Yoga Studio | Bakery | Italian Restaurant | Wine Shop | Japanese Restaurant | Gym / Fitness Center | Gym | ... | Bookstore | Grocery Store | French Restaurant | Shipping Store | Spa | Bar | Pub | Hot Dog Joint | Art Museum | Gourmet Shop |
| 2 | Central Harlem | Gym / Fitness Center | American Restaurant | Seafood Restaurant | French Restaurant | Chinese Restaurant | African Restaurant | Bar | Southern / Soul Food Restaurant | Bookstore | ... | Ethiopian Restaurant | Gym | Tapas Restaurant | Market | Event Space | Boutique | Spa | Cafeteria | Caribbean Restaurant | Library |
| 3 | Chelsea | Art Gallery | Coffee Shop | Italian Restaurant | Seafood Restaurant | American Restaurant | Ice Cream Shop | Boutique | Pizza Place | Market | ... | Bakery | Cupcake Shop | Cycle Studio | Bookstore | Park | Cheese Shop | Butcher | Club House | Nightclub | Clothing Store |
| 4 | Chinatown | Chinese Restaurant | Bakery | Cocktail Bar | Salon / Barbershop | American Restaurant | Optical Shop | Spa | Coffee Shop | Bubble Tea Shop | ... | Jewelry Store | Shanghai Restaurant | Greek Restaurant | Mexican Restaurant | Boutique | Dumpling Restaurant | Dessert Shop | Noodle House | Ice Cream Shop | Hotel |
| 5 | Civic Center | Coffee Shop | Hotel | French Restaurant | Yoga Studio | Park | Café | Gym / Fitness Center | Cocktail Bar | Spa | ... | American Restaurant | Gym | Sushi Restaurant | Wine Shop | Sporting Goods Shop | Men's Store | Building | Boxing Gym | Boutique | Burger Joint |
| 6 | Clinton | Theater | Gym / Fitness Center | Coffee Shop | Gym | Spa | Hotel | Thai Restaurant | Sandwich Place | Italian Restaurant | ... | Cocktail Bar | Wine Shop | Fried Chicken Joint | Mediterranean Restaurant | Indie Theater | Lounge | Pizza Place | Caucasian Restaurant | Roof Deck | Chinese Restaurant |
| 7 | East Harlem | Mexican Restaurant | Bakery | Latin American Restaurant | Thai Restaurant | Pizza Place | Deli / Bodega | Spa | Gas Station | Taco Place | ... | Steakhouse | French Restaurant | Spanish Restaurant | Café | Seafood Restaurant | Chinese Restaurant | Sandwich Place | Cocktail Bar | Park | Restaurant |
| 8 | East Village | Pizza Place | Cocktail Bar | Bar | Coffee Shop | Wine Bar | Japanese Restaurant | Speakeasy | Ice Cream Shop | Mexican Restaurant | ... | Italian Restaurant | Grocery Store | Gourmet Shop | Café | Ramen Restaurant | Bagel Shop | Seafood Restaurant | Greek Restaurant | Salon / Barbershop | Vietnamese Restaurant |
| 9 | Financial District | Coffee Shop | Hotel | Pizza Place | Café | Italian Restaurant | Cocktail Bar | American Restaurant | Salad Place | Falafel Restaurant | ... | Food Truck | Sandwich Place | Juice Bar | Steakhouse | Monument / Landmark | Mediterranean Restaurant | Gourmet Shop | Memorial Site | Park | Japanese Restaurant |
| 10 | Flatiron | Gym / Fitness Center | Coffee Shop | Cosmetics Shop | Café | Gym | Yoga Studio | Vegetarian / Vegan Restaurant | Furniture / Home Store | Italian Restaurant | ... | Wine Shop | Mediterranean Restaurant | Japanese Restaurant | Outdoor Sculpture | Cycle Studio | Bakery | Juice Bar | Pet Store | New American Restaurant | Toy / Game Store |
| 11 | Gramercy | Coffee Shop | Bar | Pizza Place | Ice Cream Shop | Italian Restaurant | Cocktail Bar | Grocery Store | Mexican Restaurant | Playground | ... | Park | Spa | Sandwich Place | Comedy Club | Diner | Taco Place | Wine Shop | American Restaurant | Filipino Restaurant | Gourmet Shop |
| 12 | Greenwich Village | Italian Restaurant | Café | Chinese Restaurant | Gym | Wine Bar | Coffee Shop | Jazz Club | Gourmet Shop | Mediterranean Restaurant | ... | Indian Restaurant | Pizza Place | Clothing Store | Comedy Club | Vegetarian / Vegan Restaurant | Sandwich Place | Sushi Restaurant | Record Shop | Pilates Studio | Indie Movie Theater |
| 13 | Hamilton Heights | Pizza Place | Café | Coffee Shop | Mexican Restaurant | Deli / Bodega | Yoga Studio | Cocktail Bar | Chinese Restaurant | Caribbean Restaurant | ... | School | Sushi Restaurant | Indian Restaurant | Bakery | Burger Joint | Mediterranean Restaurant | Seafood Restaurant | Liquor Store | Park | Bar |
| 14 | Hudson Yards | American Restaurant | Gym / Fitness Center | Hotel | Café | Gym | Italian Restaurant | Boat or Ferry | Park | Coffee Shop | ... | Hotel Bar | Furniture / Home Store | Comedy Club | Restaurant | Supermarket | Salad Place | Building | Stables | Cocktail Bar | Gourmet Shop |
| 15 | Inwood | Mexican Restaurant | Café | Pizza Place | Bakery | Lounge | Restaurant | Deli / Bodega | Chinese Restaurant | Wine Bar | ... | American Restaurant | Park | Moving Target | Spanish Restaurant | Bus Station | Caribbean Restaurant | Supermarket | Fast Food Restaurant | Empanada Restaurant | Farmers Market |
| 16 | Lenox Hill | Italian Restaurant | Coffee Shop | Sushi Restaurant | Pizza Place | Cocktail Bar | Gym | Café | Gym / Fitness Center | Burger Joint | ... | Bakery | Salon / Barbershop | Turkish Restaurant | Art Gallery | Playground | Cycle Studio | Wine Shop | Sporting Goods Shop | Thai Restaurant | Smoke Shop |
| 17 | Lincoln Square | Gym / Fitness Center | Plaza | Café | Italian Restaurant | Performing Arts Venue | Theater | Concert Hall | Indie Movie Theater | Wine Shop | ... | Gym | Mediterranean Restaurant | French Restaurant | Furniture / Home Store | Cycle Studio | Cosmetics Shop | Coffee Shop | Grocery Store | Bakery | Movie Theater |
| 18 | Little Italy | Chinese Restaurant | Bakery | Italian Restaurant | Café | Hotel | Spa | Thai Restaurant | Bubble Tea Shop | Cosmetics Shop | ... | Coffee Shop | Cocktail Bar | Ice Cream Shop | Salon / Barbershop | Japanese Restaurant | Pizza Place | American Restaurant | Optical Shop | Flea Market | Martial Arts Dojo |
| 19 | Lower East Side | Art Gallery | Ramen Restaurant | Café | Chinese Restaurant | Bakery | Coffee Shop | Park | Theater | Pizza Place | ... | Dance Studio | Tennis Court | Latin American Restaurant | Mediterranean Restaurant | Bubble Tea Shop | French Restaurant | Flower Shop | Filipino Restaurant | Diner | Nightclub |
| 20 | Manhattan Valley | Yoga Studio | Mexican Restaurant | Thai Restaurant | Chinese Restaurant | Pizza Place | Bar | Hawaiian Restaurant | Gym / Fitness Center | Bubble Tea Shop | ... | Café | Caribbean Restaurant | Hostel | Clothing Store | Coffee Shop | Ice Cream Shop | Playground | Korean Restaurant | Peruvian Restaurant | Dog Run |
| 21 | Manhattanville | Coffee Shop | Seafood Restaurant | Deli / Bodega | Italian Restaurant | Park | Chinese Restaurant | Mexican Restaurant | Diner | Climbing Gym | ... | Grocery Store | Bike Trail | Falafel Restaurant | Sushi Restaurant | Gastropub | Supermarket | Boutique | Bar | Fried Chicken Joint | Spanish Restaurant |
| 22 | Marble Hill | Sandwich Place | Gym | Coffee Shop | Yoga Studio | Steakhouse | Seafood Restaurant | Tennis Stadium | Supplement Shop | Bank | ... | Miscellaneous Shop | Diner | Pharmacy | Discount Store | Kids Store | Department Store | Ice Cream Shop | American Restaurant | Deli / Bodega | Pizza Place |
| 23 | Midtown | Coffee Shop | Hotel | Clothing Store | Theater | Pizza Place | Bakery | Gym | Japanese Restaurant | Spa | ... | Tailor Shop | Café | Sporting Goods Shop | Bookstore | Sandwich Place | Salon / Barbershop | French Restaurant | Steakhouse | Cycle Studio | Cuban Restaurant |
| 24 | Midtown South | Korean Restaurant | Hotel | Burger Joint | Dessert Shop | Café | Gym / Fitness Center | Japanese Restaurant | Coffee Shop | Lounge | ... | Bar | Grocery Store | New American Restaurant | Hotel Bar | American Restaurant | Cosmetics Shop | Donut Shop | Restaurant | Cycle Studio | Massage Studio |
| 25 | Morningside Heights | Park | Bookstore | American Restaurant | Coffee Shop | Deli / Bodega | Burger Joint | Ice Cream Shop | Ethiopian Restaurant | Mexican Restaurant | ... | Pizza Place | Café | New American Restaurant | Farmers Market | Seafood Restaurant | Outdoor Sculpture | Tennis Court | Sandwich Place | Salad Place | College Cafeteria |
| 26 | Murray Hill | Sandwich Place | Hotel | Bar | Burger Joint | Gym / Fitness Center | Pizza Place | Coffee Shop | Japanese Restaurant | Chinese Restaurant | ... | Cocktail Bar | Mediterranean Restaurant | French Restaurant | Sushi Restaurant | Cuban Restaurant | Jewish Restaurant | Grocery Store | Steakhouse | American Restaurant | Indian Restaurant |
| 27 | Noho | Italian Restaurant | Coffee Shop | Japanese Restaurant | Sandwich Place | Grocery Store | Pizza Place | Wine Shop | Wine Bar | Thai Restaurant | ... | Greek Restaurant | Bakery | Theater | Café | Rock Club | French Restaurant | Seafood Restaurant | Tea Room | New American Restaurant | Hotel |
| 28 | Roosevelt Island | Park | Deli / Bodega | Greek Restaurant | Coffee Shop | Bubble Tea Shop | Farmers Market | Soccer Field | Supermarket | Metro Station | ... | School | Scenic Lookout | Sandwich Place | Gym | Liquor Store | Gym / Fitness Center | Bus Line | Waterfront | Residential Building (Apartment / Condo) | Japanese Restaurant |
| 29 | Soho | Italian Restaurant | Coffee Shop | Café | Mediterranean Restaurant | Clothing Store | Bakery | Art Gallery | Hotel | Dessert Shop | ... | Salon / Barbershop | Spa | Soup Place | French Restaurant | Yoga Studio | Speakeasy | Bubble Tea Shop | Martial Arts Dojo | Garden | Boutique |
| 30 | Stuyvesant Town | Park | Bar | Coffee Shop | Gas Station | Boat or Ferry | Farmers Market | Gym | Gym / Fitness Center | Baseball Field | ... | Pet Service | Cocktail Bar | Heliport | Fountain | Dumpling Restaurant | Duty-free Shop | Food Court | Food & Drink Shop | Food | Flower Shop |
| 31 | Sutton Place | Italian Restaurant | Coffee Shop | Gym / Fitness Center | Park | Furniture / Home Store | Hotel | Pizza Place | Yoga Studio | Beer Garden | ... | Smoke Shop | Mexican Restaurant | Bagel Shop | Grocery Store | Gym | Latin American Restaurant | Health & Beauty Service | Vegetarian / Vegan Restaurant | Wine Shop | Restaurant |
| 32 | Tribeca | Park | Italian Restaurant | Wine Bar | Spa | Café | Scenic Lookout | Poke Place | Playground | Coffee Shop | ... | Basketball Court | Men's Store | Art Gallery | Bakery | Greek Restaurant | Skate Park | American Restaurant | Wine Shop | Hotel | Seafood Restaurant |
| 33 | Tudor City | Park | Café | Mexican Restaurant | Deli / Bodega | Diner | Greek Restaurant | Dog Run | Sushi Restaurant | Gym / Fitness Center | ... | Thai Restaurant | Pizza Place | Restaurant | Garden | Coffee Shop | Salon / Barbershop | Salad Place | Shanghai Restaurant | Burger Joint | Spa |
| 34 | Turtle Bay | Coffee Shop | Italian Restaurant | Café | Park | Deli / Bodega | Hotel | Sushi Restaurant | French Restaurant | Seafood Restaurant | ... | Thai Restaurant | Diner | Plaza | Garden | Dog Run | Spa | Pub | Japanese Restaurant | Karaoke Bar | Turkish Restaurant |
| 35 | Upper East Side | Italian Restaurant | Coffee Shop | Exhibit | Bakery | Yoga Studio | Gym / Fitness Center | Pizza Place | French Restaurant | Juice Bar | ... | Wine Shop | Hotel | Sushi Restaurant | Salad Place | American Restaurant | Art Gallery | Grocery Store | Park | Plaza | History Museum |
| 36 | Upper West Side | Italian Restaurant | Bakery | Wine Bar | Coffee Shop | Bar | American Restaurant | Ice Cream Shop | Pizza Place | Bagel Shop | ... | Bookstore | Seafood Restaurant | Middle Eastern Restaurant | Thai Restaurant | Drugstore | Food & Drink Shop | Ramen Restaurant | Sushi Restaurant | Garden | Salad Place |
| 37 | Washington Heights | Café | Bakery | Grocery Store | Mobile Phone Shop | Chinese Restaurant | Supermarket | Tapas Restaurant | Bank | Mexican Restaurant | ... | Supplement Shop | Pizza Place | Sandwich Place | Gym | Deli / Bodega | New American Restaurant | Spa | Wine Shop | Latin American Restaurant | Donut Shop |
| 38 | West Village | Italian Restaurant | Wine Bar | Park | New American Restaurant | American Restaurant | Coffee Shop | Bakery | Seafood Restaurant | Jazz Club | ... | Board Shop | Food & Drink Shop | Cycle Studio | Gym | Pizza Place | Cocktail Bar | Steakhouse | Ice Cream Shop | Sandwich Place | Cheese Shop |
| 39 | Yorkville | Italian Restaurant | Coffee Shop | Gym | Bar | Deli / Bodega | Sushi Restaurant | Wine Shop | Mexican Restaurant | Diner | ... | Japanese Restaurant | Bagel Shop | Chinese Restaurant | Sandwich Place | Park | Ice Cream Shop | Vietnamese Restaurant | Pharmacy | Butcher | German Restaurant |

**Figure 3.5**

# 4. Methodology - Predictive Modelling:

As this is a typical clustering algorithm, we choose to implement this on partition based clustering which fits exactly to the problem definition.

## 4.1 Partition Based Clustering:

It is a clustering methods used to classify observations, within a data set, into multiple groups based on their similarity.

**K-Means Algorithm**:

K-means clustering is a partitioning method and as anticipated, this method decomposes a dataset into a set of disjoint clusters. Given a dataset, a partitioning method constructs several partitions of this data, with each partition representing a cluster.

So we convert the above top common Venues table into machine readable format, by converting Venue names with the mean of their occurrences with added weight of how often the Venues are checked in and/or ratings of the Venues, if there is any such preference of the user for ratings.

You can see the sample data below (**Figure 4.1**),

| American Restaurant | Art Gallery | Asian Restaurant | BBQ Joint | Bakery | Beer Bar | Belgian Restaurant | Bistro | Bookstore | Breakfast Spot | ... | Video Game Store | Video Store | Vietnamese Restaurant | Volleyball Court | Waterfront | Whisky Bar | Wine Shop | Wings Joint | Women's Store | Yoga Studio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.037500 | 0.012500 | 0.012500 | 0.012500 | 0.012500 | 0.012500 | 0.0125 | 0.012500 | 0.012500 | 0.012500 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.000000 |
| 0.016667 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 | 0.016667 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.033333 | 0.000 | 0.000000 | 0.000000 |
| 0.011628 | 0.000000 | 0.000000 | 0.000000 | 0.034884 | 0.000000 | 0.0000 | 0.000000 | 0.034884 | 0.000000 | ... | 0.000000 | 0.00 | 0.011628 | 0.000000 | 0.00 | 0.000000 | 0.034884 | 0.000 | 0.000000 | 0.034884 |
| 0.047619 | 0.023810 | 0.000000 | 0.023810 | 0.000000 | 0.023810 | 0.0000 | 0.000000 | 0.023810 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.000000 |
| 0.030000 | 0.120000 | 0.000000 | 0.000000 | 0.020000 | 0.010000 | 0.0000 | 0.000000 | 0.020000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.010000 | 0.000 | 0.010000 | 0.000000 |
| 0.030000 | 0.000000 | 0.020000 | 0.000000 | 0.060000 | 0.000000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.010000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.010000 |
| 0.022989 | 0.000000 | 0.011494 | 0.000000 | 0.011494 | 0.000000 | 0.0000 | 0.000000 | 0.011494 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.011494 | 0.022989 | 0.000 | 0.000000 | 0.034483 |
| 0.030000 | 0.010000 | 0.000000 | 0.000000 | 0.010000 | 0.000000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.030000 | 0.000 | 0.000000 | 0.000000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.095238 | 0.023810 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.000000 |
| 0.010000 | 0.010000 | 0.000000 | 0.000000 | 0.010000 | 0.010000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.020000 | 0.000000 | 0.00 | 0.000000 | 0.010000 | 0.000 | 0.000000 | 0.000000 |
| 0.030000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.0000 | 0.000000 | 0.000000 | 0.010000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.010000 | 0.000 | 0.000000 | 0.010000 |
| 0.010000 | 0.010000 | 0.000000 | 0.000000 | 0.020000 | 0.000000 | 0.0000 | 0.000000 | 0.010000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.030000 | 0.000 | 0.000000 | 0.030000 |
| 0.027778 | 0.013889 | 0.000000 | 0.000000 | 0.000000 | 0.013889 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.027778 | 0.000 | 0.000000 | 0.013889 |
| 0.010000 | 0.010000 | 0.000000 | 0.000000 | 0.010000 | 0.010000 | 0.0000 | 0.000000 | 0.000000 | 0.010000 | ... | 0.000000 | 0.00 | 0.010000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.010000 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.033898 | 0.000000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000 | 0.000000 | 0.033898 |
| 0.054545 | 0.000000 | 0.000000 | 0.018182 | 0.000000 | 0.000000 | 0.0000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.00 | 0.000000 | 0.018182 | 0.000 | 0.000000 | 0.000000 |

**Figure 4.1**

After optimising the input for the algorithm, we have grouped the Neighbourhoods into different clusters as seen below (**Figure 4.2**),


**Figure 4.2**

As you can see there are many clusters built based on the venue data that we have fetched from the Foursquare API, if we match the cluster from the source city, we can eliminate all other clusters which will enhance our visual analysing capabilities.

# 5. Results

By eliminating all other clusters which does not match our source Neighbourhood, we will end up having the below map (**Figure 5.1**), where all the red coloured clusters are not matched and the green coloured clusters are our similar Neighbourhoods (Right side) when compared to the user's source Neighbourhood (Left Side).


**Figure 5.1**

As you can see in the Right side of the map above, it also has a pin representing the user's new work location based on the provided address.

Using the Haversine Formula to find the distance between the New Work Location and the suggested similar neighbourhoods, below (**Figure 5.2**) are the distances accordingly,

```
-------- Tribeca --------
Overall Similar Venues: 11 / 50
Distance from new work place:  3.86  km
-------- West Village --------
Overall Similar Venues: 25 / 50
Distance from new work place:  2.52  km
-------- Turtle Bay --------
Overall Similar Venues: 24 / 50
Distance from new work place:  1.42  km
-------- Tudor City --------
Overall Similar Venues: 17 / 50
Distance from new work place:  1.15  km
```

**Figure 5.2**

We can visualise the top 20 common venues across the source Neighbourhood and Destination Neighbourhood using the bar chart, and below are the series of charts which shows the similarities among them,

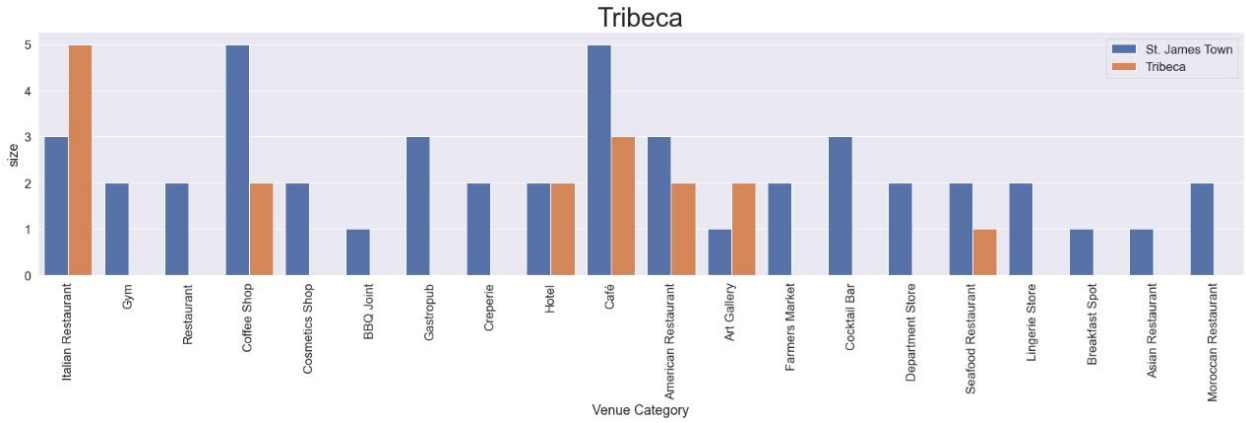


**Figure 5.3(a)**

Tribeca matches only 7 top similar venues and only 11 overall similar Venues. Lets see how other Neighbourhoods are doing.

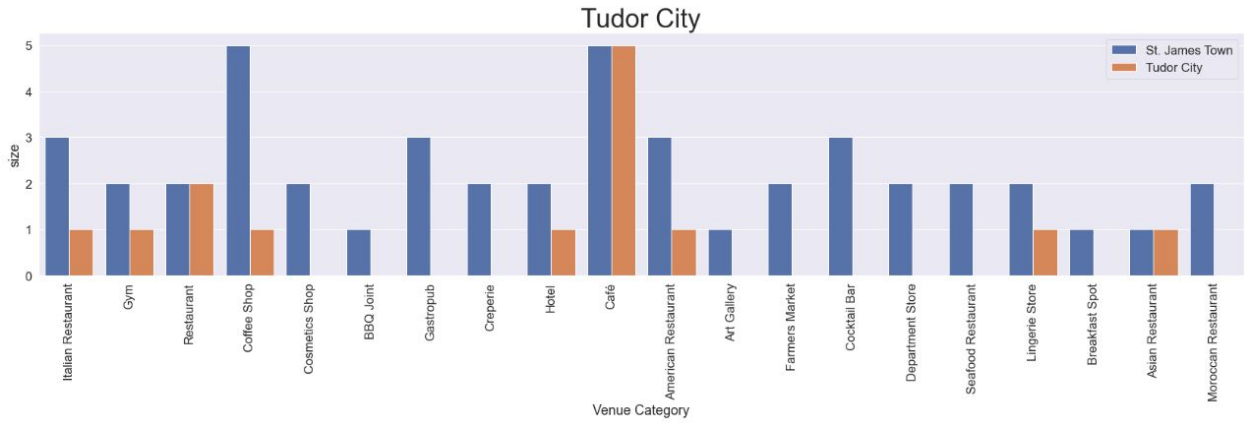


**Figure 5.3(b)**

Statistics seem to be improving, Tudor City has 9 similar top common Venues and 17 Overall similar Venues.

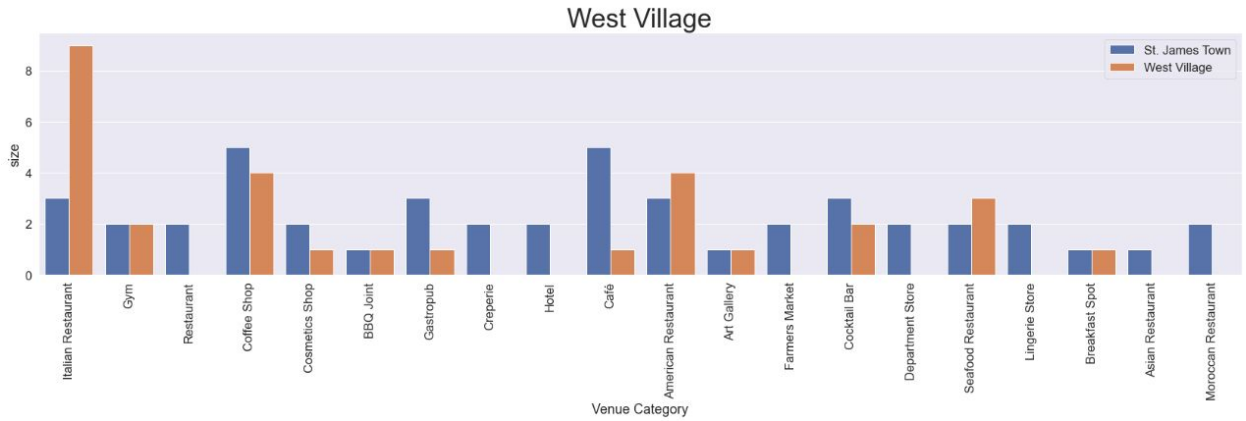


**Figure 5.3(c)**

Having 12 similar top common and 25 overall similar Venues West Village can be considered for recommendation.

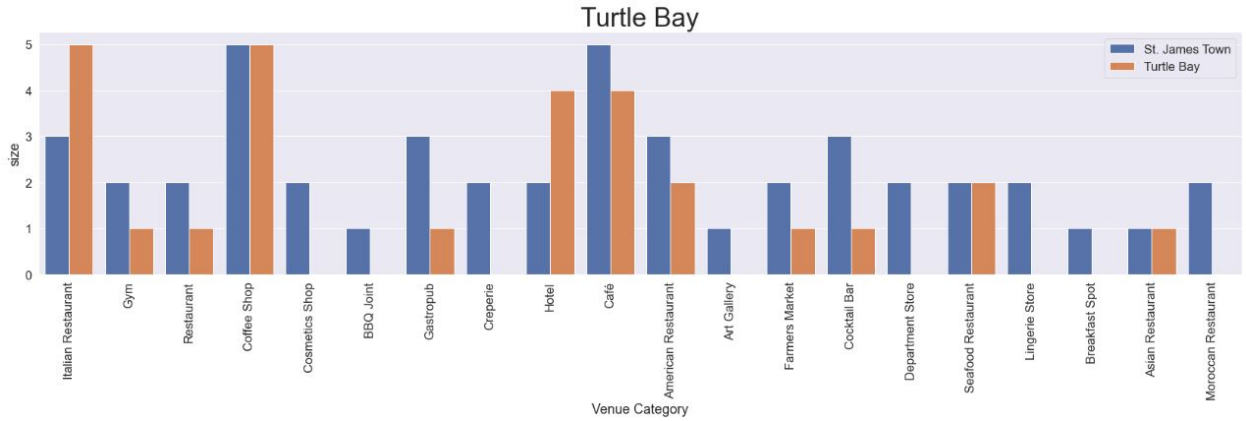


**Figure 5.3(d)**

Woohoo! We have a Battle of Neighbourhoods. Turtle Bay having 12 similar top common Venues competes with West Village for the recommendation.

# 6. Discussion

## 6.1 Recommendation

As we saw in the visualizations above, Compared to the user's current neighborhood, Turtle Bay and West Village are the suggested Neighborhoods based on the top common venue preferences.

*Turtle Bay Neighbourhood* has 12 top common similar venues which includes different genres of restaurants like Italian, American, Seafood and Asian. And even has a Gastro pub and a Cocktail bar if the user prefers Alcohol. The Neighborhood has a Gym for some quality time with fitness. Also has a Farmer's Market if the user prefers picking up the veggies to cook some nutritious food. And there are no shortages of Coffee shops and Cafe to spend quality time with friends or just have a productive time with themselves.

This Neighborhood is 1.15 KM far from the user's new work place, which is a short walk.

*West Village Neighbourhood* also has 12 top common similar venues. Even this neighbourhood has many restaurants -- 9 Italian, 4 American, 3 Seafood Restaurants, BBQ Joint and a breakfast spot. For alcohol lovers, there is 1 Gastro pub and 2 Cocktail bars. There are 2 Gyms in the Neighborhood, so get ready to get fit. There are coffee shops and cafes available to be productive or just for fun. Also there is a cosmetic shop in the block, if you prefer to have some makeover.

This Neighbourhood is 2.52 KM far from the user's new work place, cycling is both effective towards health and time.

### 6.2 Observations

As I used Foursquare API to get the data on Venues across neighbourhoods, the data is constantly changing as the check-in timings of people changes often and you cannot expect the mindset of the people will remain the same when you cross the borders. The more consistent and reliable the data is, the quality of prediction will be higher. The prediction can also be made better if we know the user's requirement / preferences of Venues.

I have used the K-Means algorithm for clustering the neighbourhoods. When I used the Elbow method on the data of Venues as explained before, the best k value was not giving me enough clue. You can see the screenshot (**Figure 6.1**).

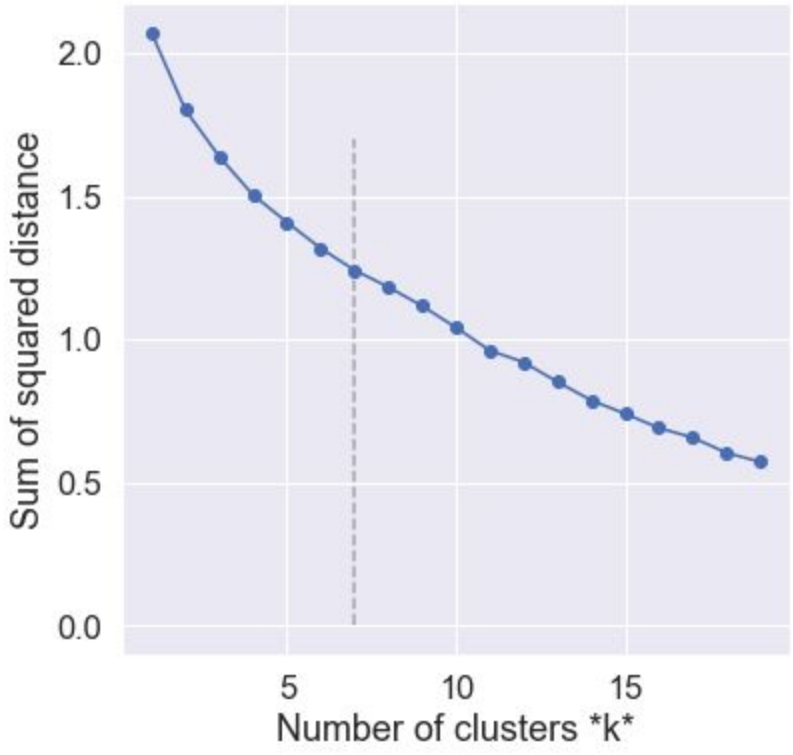


**Figure 6.1**

That might be because the dimensions of data used is not clear enough to cluster them. For instance, each neighbourhood has one or the other similar venues among them but the difference between few neighbourhoods are very minimal.

## 7. Conclusion

As we saw the prediction of similar cities comparing source and destination venues and their information. The result is purely dependent on the active venues and will often change on every execution as active check-ins of venues in different neighbourhoods change every day.

This system is helpful for not only the people who want to relocate to another city, but also for the people who want to open a new Venue like restaurants, Pubs, Cafe's, etc.., in a similar Neighbourhood.

## 7. Future directions

Fetching the data on crime rates, pollution, cost of living for each Neighborhood will enhance the result with better quality. Collecting the data for the same from different sources of the web, validating the integrity of the data is a challenge. Also planning to clean up the input dimensions to K-Means for selecting the best “k” value as discussed in the Discussion section.

Also looking forward to automating the same system to take the input for source and destination Neighbourhoods and suggest similar places to visit or relocate to.

## 8. References


https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

https://cocl.us/new_york_dataset/newyork_data.json

https://pandas.pydata.org/pandas-docs/stable/reference/api/

https://medium.com/@petehouston/calculate-distance-of-two-locations-on-earth-using-python-1501b1944d97

https://stats.stackexchange.com/questions/89679/what-do-you-do-when-theres-no-elbow-point-for-kmeans-clustering