# Battle of Neighborhoods

The Shop Explorer

# Business Problem

Vancouver is a major city in western Canada and is the most populous city in the province of British Columbia. Being the third largest metropoliton city, it has the largest population density in Canada. The basic necessities of such a large number of people is satisfied by the numerous shopping venues present accross the huge metropolis.

This project aims at segmenting the shopping venues into clusters based on their distance from the neighborhood centre they are present and explore various aspects like most common distances for shops to be found, etc. Intuitively, based on geography of the city, the shops that are located within a close range to any neighborhood gets grouped into one cluster. This helps in identifying shops that are outliers and customers may need to go out of their way or through a secluded area to have access to them. This can result in possible threats like robbing, lack of timely emergency services, or in the worst case scenarios life threatening events.

# Data Description

## Wikipedia Data:

The data is fetched as html text from wikipedia and gives the Postal Code, City Name, and Neighborhood Name for the Metropolitan city of Vancouver. web scraping and cleaning is used to scrap the data and fetch the table having the required data fields and ultimately convert it into a pandas DataFrame.

Link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_V

# Data Description

## Geocoder Data

Further data for Longitudes and Latitudes is fetched using Geocoder library.

# Data Description

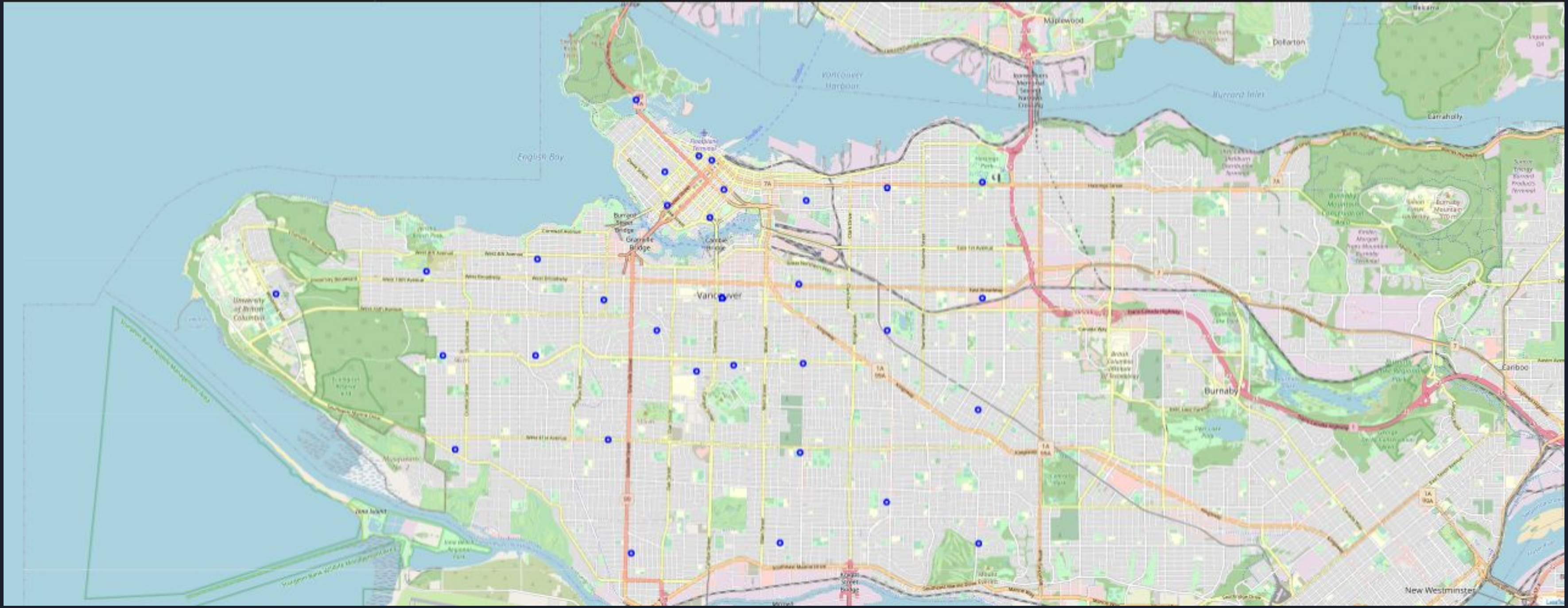
## Foursquare API Data

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about shops inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 15 km. The data retrieved from Foursquare contained information of shops within a specified distance of the longitude and latitude of the postcodes. The information obtained per shop is as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Shop
5. Name of the Shop/Mall
6. Shop Latitude
7. Shop Longitude
8. Shop Distance from center of Neighborhood
9. Shop Address
10. Shop Category

# Methodology

## Map of Vancouver with all Neighborhoods



# Methodology

## FourSquare API

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

## Work Flow

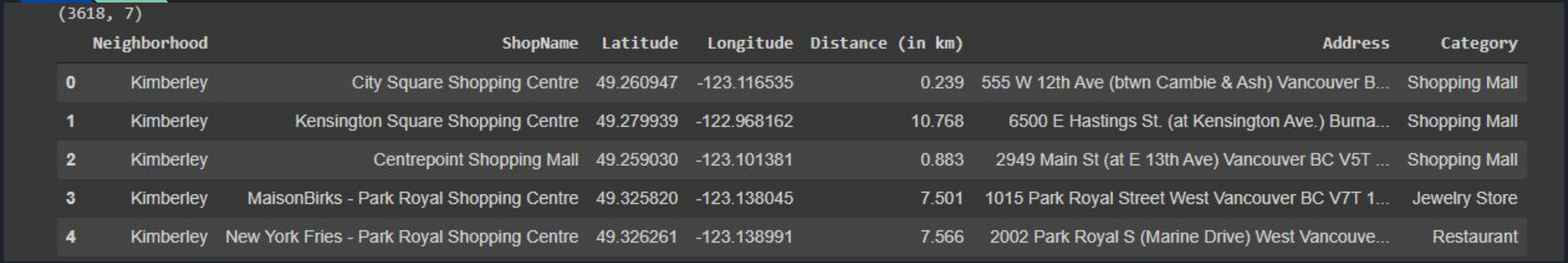
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 50 and the radius parameter would be set to 15 km.

## Clustering Approach

To achieve the project goal, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like Vancouver. To be able to do that, we need to cluster data using a form of unsupervised machine learning: k-means clustering algorithm

# Results

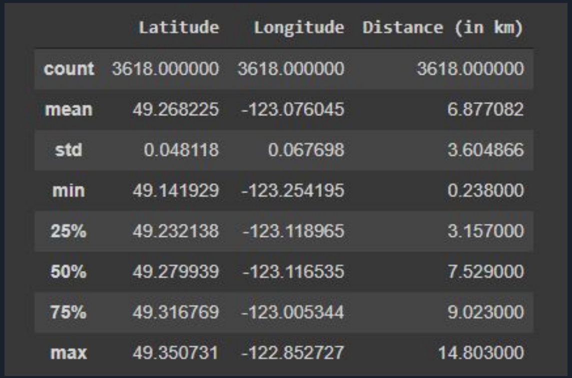
## Example of DataFrame after retrieving information from FourSquare:



```
(3618, 7)
```

| | Neighborhood | ShopName | Latitude | Longitude | Distance (in km) | Address | Category |
|---|---|---|---|---|---|---|---|
| 0 | Kimberley | City Square Shopping Centre | 49.260947 | -123.116535 | 0.239 | 555 W 12th Ave (btwn Cambie & Ash) Vancouver B... | Shopping Mall |
| 1 | Kimberley | Kensington Square Shopping Centre | 49.279939 | -122.968162 | 10.768 | 6500 E Hastings St. (at Kensington Ave.) Burna... | Shopping Mall |
| 2 | Kimberley | Centrepoint Shopping Mall | 49.259030 | -123.101381 | 0.883 | 2949 Main St (at E 13th Ave) Vancouver BC V5T ... | Shopping Mall |
| 3 | Kimberley | MaisonBirks - Park Royal Shopping Centre | 49.325820 | -123.138045 | 7.501 | 1015 Park Royal Street West Vancouver BC V7T 1... | Jewelry Store |
| 4 | Kimberley | New York Fries - Park Royal Shopping Centre | 49.326261 | -123.138991 | 7.566 | 2002 Park Royal S (Marine Drive) West Vancouve... | Restaurant |

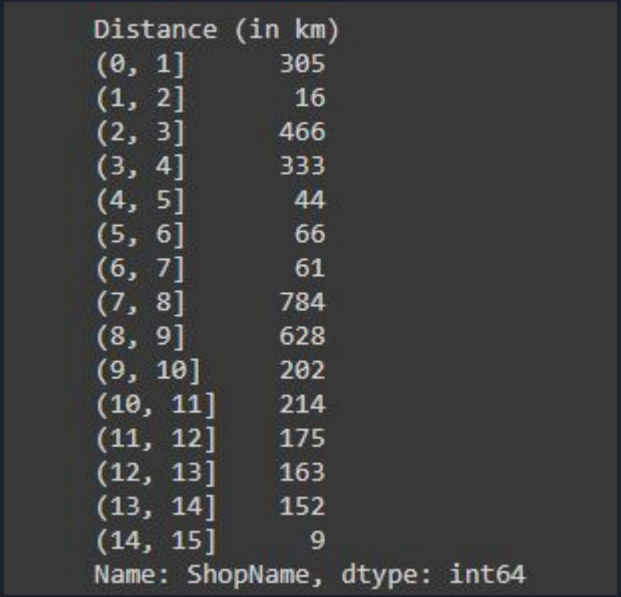
## Description of shops:



| | Latitude | Longitude | Distance (in km) |
|---|---|---|---|
| count | 3618.000000 | 3618.000000 | 3618.000000 |
| mean | 49.268225 | -123.076045 | 6.877082 |
| std | 0.048118 | 0.067698 | 3.604866 |
| min | 49.141929 | -123.254195 | 0.238000 |
| 25% | 49.232138 | -123.118965 | 3.157000 |
| 50% | 49.279939 | -123.116535 | 7.529000 |
| 75% | 49.316769 | -123.005344 | 9.023000 |
| max | 49.350731 | -122.852727 | 14.803000 |

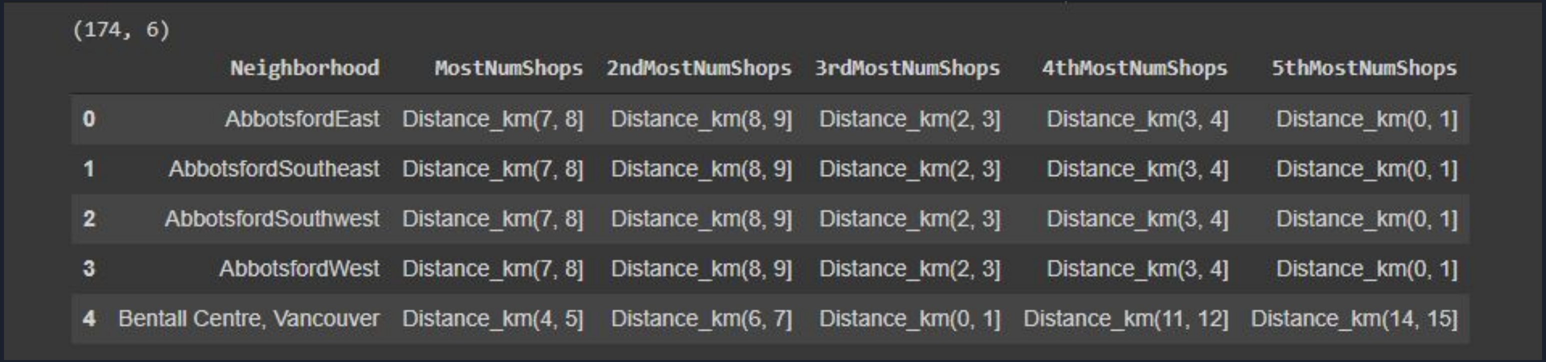
# Results

Number of Shops within various distance ranges:



```
Distance (in km)
(0, 1]       305
(1, 2]        16
(2, 3]       466
(3, 4]       333
(4, 5]        44
(5, 6]        66
(6, 7]        61
(7, 8]       784
(8, 9]       628
(9, 10]      202
(10, 11]     214
(11, 12]     175
(12, 13]     163
(13, 14]     152
(14, 15]       9
Name: ShopName, dtype: int64
```

Distance range where most number of shops are present for each neighborhood:

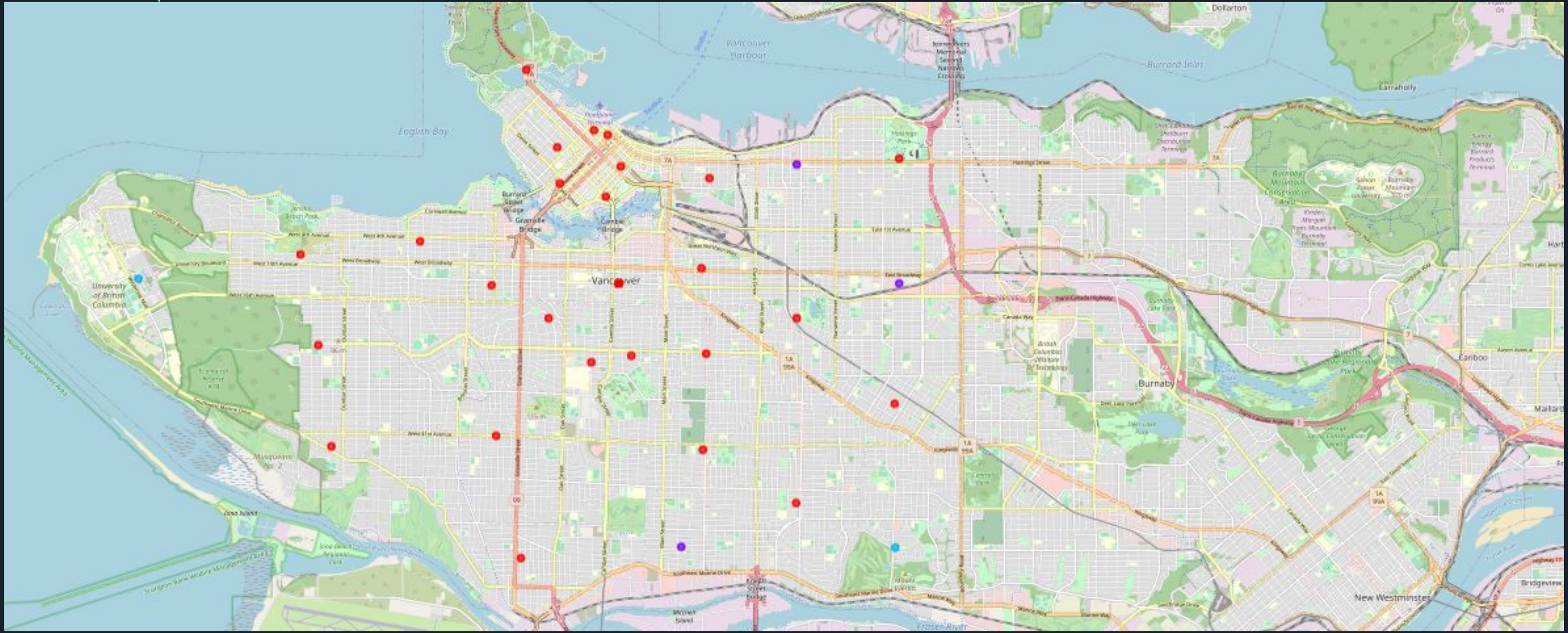


```
(174, 6)
```

| | Neighborhood | MostNumShops | 2ndMostNumShops | 3rdMostNumShops | 4thMostNumShops | 5thMostNumShops |
|---|---|---|---|---|---|---|
| 0 | AbbotsfordEast | Distance_km(7, 8] | Distance_km(8, 9] | Distance_km(2, 3] | Distance_km(3, 4] | Distance_km(0, 1] |
| 1 | AbbotsfordSoutheast | Distance_km(7, 8] | Distance_km(8, 9] | Distance_km(2, 3] | Distance_km(3, 4] | Distance_km(0, 1] |
| 2 | AbbotsfordSouthwest | Distance_km(7, 8] | Distance_km(8, 9] | Distance_km(2, 3] | Distance_km(3, 4] | Distance_km(0, 1] |
| 3 | AbbotsfordWest | Distance_km(7, 8] | Distance_km(8, 9] | Distance_km(2, 3] | Distance_km(3, 4] | Distance_km(0, 1] |
| 4 | Bentall Centre, Vancouver | Distance_km(4, 5] | Distance_km(6, 7] | Distance_km(0, 1] | Distance_km(11, 12] | Distance_km(14, 15] |

# Results

## Final Clustering Result:



# Libraries Used

- Pandas: For creating and manipulating dataframes.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Scikit Learn: For importing k-means clustering.
- Geocoder: To retrieve Location Data.
- Beautiful Soup and Requests: To scrap and library to handle http requests.
- Matplotlib: Python Plotting Module.