**Applied Data Science Capstone**

**~Battle of The Neighborhoods ~**

Dorota Buczynska

June 2, 2019

1. **Introduction**

1.1 Business Problem:

As a busy worker who would like to move to Chicago I need to find out which neighborhoods are replete with restaurants and grocery stores in close proximity. I need to figure out top three clusters of neighbrhoods.

1.2 Interest:

Busy office workers who have no time to cook at home and prefer dining at the restaurants.

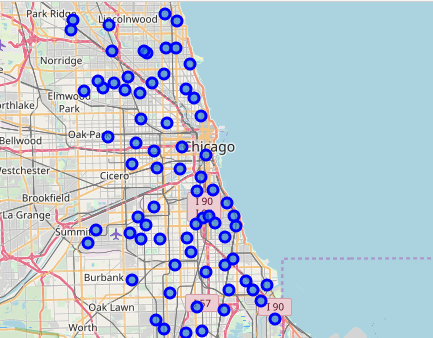
1. **Data acquisition**

**2.1** Postcode, Borough and geo coordinates of Chicago are retrieved from <https://data.cityofchicago.org> json file while boroughs divided by the neighborhoods from Wikipedia webpage <https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago>.

BeautifulSoup is used to do scraping. The data is merged and grouped by unique borough and postal code.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **PostalCodes** | **Borough** | **Latitude** | **Longitude** | **Neighborhood** |
| **0** | 60605 | Near South Side | 41.866856 | -87.628651 | Central Station, Dearborn Park, Museum Campus,... |
| **1** | 60607 | Near West Side | 41.874327 | -87.660976 | Fulton River District, Greektown, Illinois Med... |
| **2** | 60608 | Lower West Side | 41.852673 | -87.663769 | East Pilsen, Heart of Chicago, Lower West Side... |
| **3** | 60609 | Bridgeport | 41.829990 | -87.641066 | Bridgeport |
| **4** | 60609 | Fuller Park | 41.801699 | -87.632974 | Fuller Park |
| **5** | 60609 | New City | 41.795649 | -87.642615 | Back of the Yards, Canaryville, New City |
| **6** | 60610 | Near North Side | 41.906573 | -87.635514 | Cabrini–Green, Gold Coast, Goose Island, Magni... |
| **7** | 60612 | East Garfield Park | 41.870912 | -87.699887 | East Garfield Park, Fifth City |
| **8** | 60613 | Uptown | 41.959978 | -87.649712 | Buena Park, Clarendon Park, Margate Park, New ... |
| **9** | 60614 | Lincoln Park | 41.924497 | -87.644522 | Lincoln Park, Old Town Triangle, Park West, Ra... |
| **10** | 60615 | Grand Boulevard | 41.804564 | -87.624484 | Grand Boulevard, Legends South (Robert Taylor ... |

Final result of merged data is the map of Chicago where each dot represents a bourough:



**2.2** Venue information is retrieved from FourSquare

The code retrieves venues based on 500 meters radius and a limit of 300 objects from the boroughs’ centers.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Neighborhood Latitude** | **Neighborhood Longitude** | **Venue** | **Venue Latitude** | **Venue Longitude** | **Venue Category** |
| **0** | Central Station, Dearborn Park, Museum Campus,... | 41.866856 | -87.628651 | Stan's Donuts & Coffee | 41.867516 | -87.626402 | Coffee Shop |
| **1** | Central Station, Dearborn Park, Museum Campus,... | 41.866856 | -87.628651 | Tejas Yoga | 41.865584 | -87.626084 | Yoga Studio |
| **2** | Central Station, Dearborn Park, Museum Campus,... | 41.866856 | -87.628651 | Trader Joe's | 41.867820 | -87.625434 | Grocery Store |
| **3** | Central Station, Dearborn Park, Museum Campus,... | 41.866856 | -87.628651 | Kai Sushi | 41.865916 | -87.627234 | Sushi Restaurant |
| **4** | Fulton River District, Greektown, Illinois Med... | 41.874327 | -87.660976 | Jade Court Chinese Cuisine | 41.873387 | -87.657045 | Chinese Restaurant |
| **5** | Fulton River District, Greektown, Illinois Med... | 41.874327 | -87.660976 | Hubbard Street Dance Center | 41.877446 | -87.656651 | Dance Studio |
| **6** | Fulton River District, Greektown, Illinois Med... | 41.874327 | -87.660976 | CTA - Racine | 41.876490 | -87.658770 | Train Station |

**2.3** There are two ratios calculated: for Restaurants among all venue categories for all the neighborhoods and ratios for grocery stores since they are used as a factor in k-means.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **Neighborhood Latitude** | **Neighborhood Longitude** | **Restaurants Count** | **Grocery Count** | **Ratio** | **Ratio\_grocery** |
| **0** | Albany Park, Mayfair, North Mayfair, Ravenswoo... | 41.971143 | -87.709627 | 9.0 | 2.0 | 0.281250 | 0.062500 |
| **1** | Altgeld Gardens, Eden Green, Golden Gate, Rive... | 41.657432 | -87.606496 | 0.0 | 1.0 | 0.000000 | 0.250000 |
| **2** | Andersonville, Edgewater, Edgewater Beach, Edg... | 41.976216 | -87.669762 | 15.0 | 1.0 | 0.180723 | 0.012048 |
| **3** | Archer Heights | 41.803046 | -87.722007 | 3.0 | 0.0 | 0.150000 | 0.000000 |
| **4** | Armour Square, Chinatown, Wentworth Gardens | 41.843952 | -87.635318 | 4.0 | 0.0 | 0.333333 | 0.000000 |

**2.4** The other factor for the k-means algorithm is the Top 5 venue categories for each neighborhood.

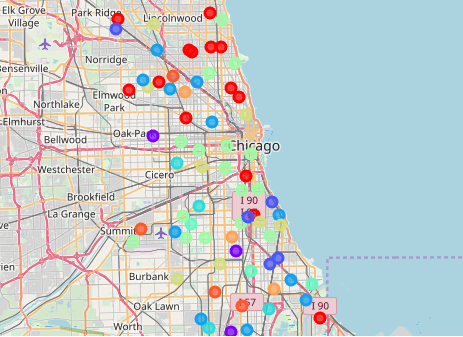
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Neighborhood** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| **0** | Albany Park, Mayfair, North Mayfair, Ravenswoo... | Restaurant | Ice Cream Shop | Grocery | Park | Bank |
| **1** | Altgeld Gardens, Eden Green, Golden Gate, Rive... | Food | Grocery | Park | Clothing Store | Yoga Studio |
| **2** | Andersonville, Edgewater, Edgewater Beach, Edg... | Restaurant | Sandwich Place | Pet Store | Salon / Barbershop | Bookstore |
| **3** | Archer Heights | Restaurant | Pizza Place | Discount Store | Sandwich Place | Cosmetics Shop |
| **4** | Armour Square, Chinatown, Wentworth Gardens | Restaurant | Pizza Place | Park | Breakfast Spot | Flower Shop |

**2.5** The data frames from 2.3 and 2.4 were merged to retrieve the final dataframe for the k-means algorithm. Because the k-means can’t be applied to strings, the categories were converted to numbers.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Neighborhood** | **Neighborhood Latitude** | **Neighborhood Longitude** | **Restaurants Count** | **Grocery Count** | **Ratio** | **Ratio\_grocery** | **1st Most Common Venue** | **2nd Most Common Venue** | **3rd Most Common Venue** | **4th Most Common Venue** | **5th Most Common Venue** |
| **Albany Park, Mayfair, North Mayfair, Ravenswood Manor** | 41.971143 | -87.709627 | 9.0 | 2.0 | 0.281250 | 0.062500 | Restaurant | Ice Cream Shop | Grocery | Park | Bank |
| **Altgeld Gardens, Eden Green, Golden Gate, Riverdale** | 41.657432 | -87.606496 | 0.0 | 1.0 | 0.000000 | 0.250000 | Food | Grocery | Park | Clothing Store | Yoga Studio |
| **Andersonville, Edgewater, Edgewater Beach, Edgewater Glen, Lakewood / Balmoral** | 41.976216 | -87.669762 | 15.0 | 1.0 | 0.180723 | 0.012048 | Restaurant | Sandwich Place | Pet Store | Salon / Barbershop | Bookstore |
| **Archer Heights** | 41.803046 | -87.722007 | 3.0 | 0.0 | 0.150000 | 0.000000 | Restaurant | Pizza Place | Discount Store | Sandwich Place | Cosmetics Shop |
| **Armour Square, Chinatown, Wentworth Gardens** | 41.843952 | -87.635318 | 4.0 | 0.0 | 0.333333 | 0.000000 | Restaurant | Pizza Place | Park | Breakfast Spot | Flower Shop |

1. **Methodology**

Used k-means algorithm to cluster the neighborhoods based on 1) ratios of Restaurants and grocery stores among all venue categories, and 2) top 5 venue categories in each neighborhood.



1. **Result**

The neighborhoods in Chicago are clustered into 10 clusters.

Best Cluster is 0 (red circles); it has the highest ratios of restaurants and grociery stores and the biggest # of top 5 venue categories. So these neighborhoods are recommended.

Second Best Group is 3 (ocean blue); Third Best Group is 7 (olive);