

Finding optimal locations for an Adult Day Care Center in Queens

1, Problem

Problem In this project we will try to find an optimal location for an adult day care center in the Queens Borough of New York City whose main focus will be on recreational activities and social stimulation for the elderly population who would otherwise stay at home alone. The recreational activities would include: daily exercise regimes in a local park (tai chi, yoga, pilates, walking), while the social stimulation would consist of arts and crafts, music, games (bingo, scrabble, etc.) and general socialization and conversations to form friendly relationships. Only in case of a bad weather daily exercise regimes would be performed indoor. The center would have a nurse on-site so that participants' vital signs can be checked and evaluated regularly. The center would also fill prescriptions at a local pharmacy if participants request such service. In addition, the center would provide healthy meals and snacks and transportation to participants. Since the center would provide daily exercise regimes, we prefer locations as close to parks as possible. To avoid competition, we don't want to be in a proximity of existing adult day care centers. To be able to fill prescriptions we want to be in a proximity of a pharmacy.

Specifically, since we want to be as close to parks as possible our problem will be to identify parks in Queens satisfying the following 2 conditions:

1. No existing adult day care center within 1 km of the park latitude and longitude coordinates.
2. At least one pharmacy within 2 km of the park latitude and longitude coordinates.

2, Analysis

Basic explanatory data analysis shows that we have good data. None of our datasets:

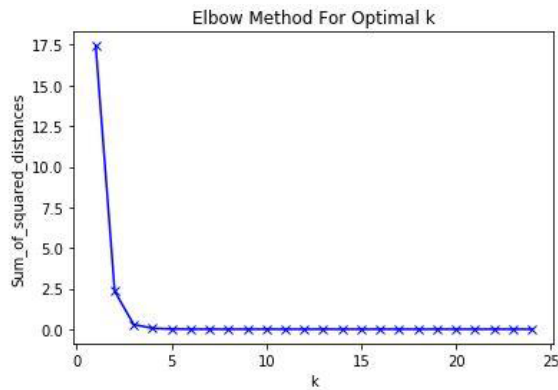
1. queens_neighborhoods (neighborhoods data for Queens)
2. queens_pharmacies_and_parks (parks and pharmacies venues data in Queens)
3. queens_adult_care_services_data (adult day care service data in Queens)

has any null or NaN value. There are 81 neighborhoods in Queens, 78 pharmacies and parks, and quite a lot of adult day care service centers (121). Summary Statistics of 4

numeric columns show data consistency for all 3 datasets. For all 3 datasets mean values of Latitude and Longitude are almost equal to the geographical coordinates of Queens: 40.6524927, -73.7914214158161. Top 3 venues in Queens are Pizza Place, Deli/Bodega, and Chinese Restaurant.

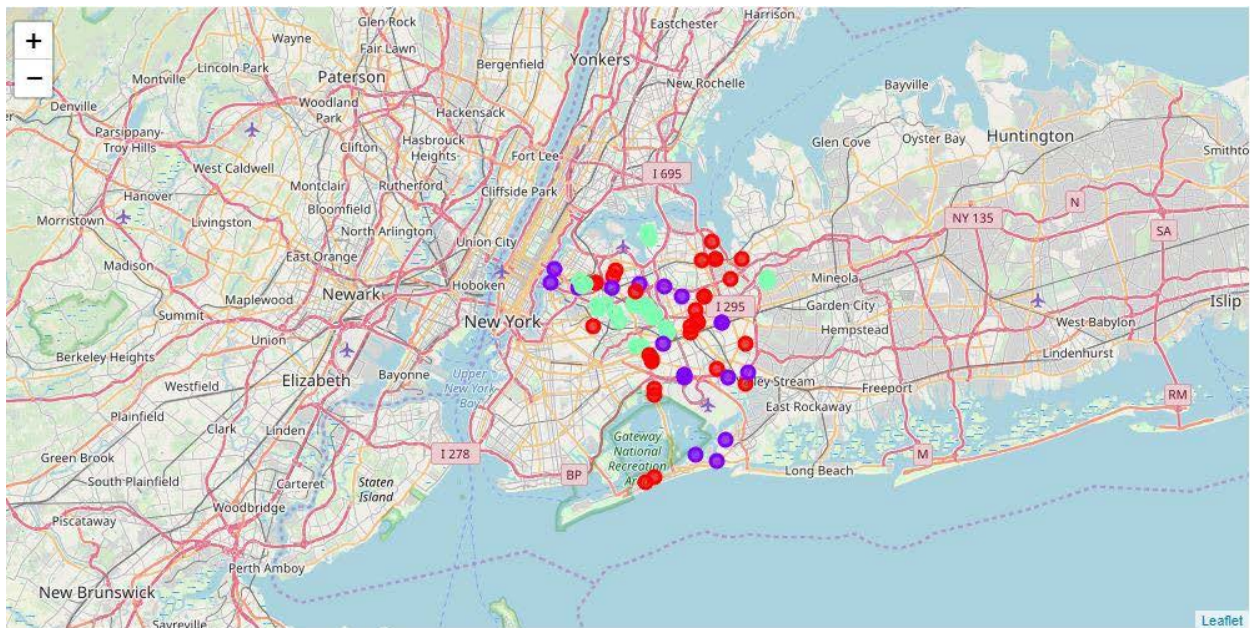
Having established that we have good data, in order to cluster pharmacies and parks in Queens using k-means clustering method, we need to do one hot encoding of the "Venue Category" categorical variable for the queens_pharmacies_and_parks dataset. This variable has only 2 values: Park or Pharmacy. One hot encoding is a process by which categorical variables are converted into a

form that could be provided to Machine learning algorithms to do a better job in prediction. By applying the “elbow” method to determine k for k-means clustering we find that the optimal k for the queens_pharmacies_and_parks dataset is 3 as can be seen from the below plot:



In the plot above the elbow is at k=3 indicating that the optimal k for the queens_pharmacies_and_parks_grouped dataset is 3

We generate 3 clusters: one with parks only (violet color), one with pharmacies only (red color), and one with pharmacies and parks mixed together (green color) as can be seen from the below folium map:



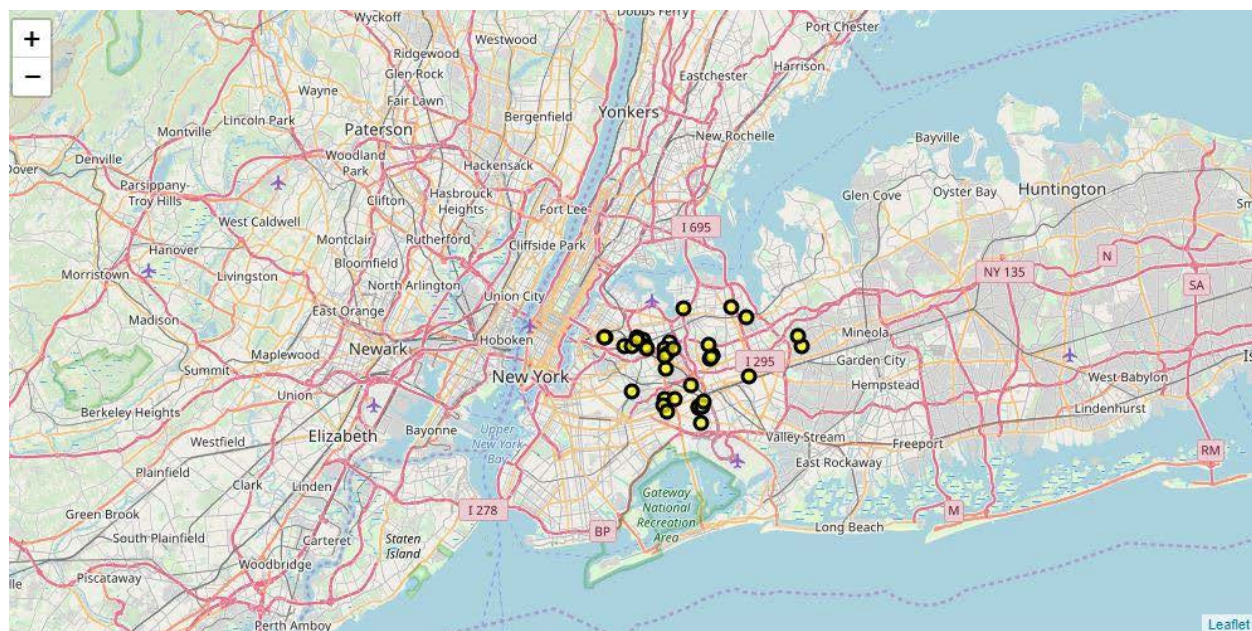
The analysis shows that some of the neighborhood names (NTA column) are slightly different in this dataset as compared to the queens_pharmacies_and_parks dataset, for example **Elmhurst-Maspeth** in the queens_adult_care_services_data vs. **Elmhurst** in the queens_pharmacies_and_parks dataset. We decided to split the NTA column on "-" and take the first part of the split as the neighborhood name.

With this data transformation in place we joined the queens_adult_care_services_data with the queens_pharmacies_and_parks data enriched with respective cluster labels using the neighborhood name as the inner join column in order to see the distribution of respective adult day care service centers across clusters. 15 neighborhoods were not joined

```
Now check which neighborhoods from queens_adult_care_serv_neigh_list are not in the queens_pharmacies_and_parks_neigh_list

queens_adult_care_serv_neigh_list = sorted(queens_adult_care_services_message['neighborhood']).unique().tolist()
p = set(queens_pharmacies_and_parks_neigh_list)
queens_adult_care_serv_neigh_not_in_ph_and_park_list = [x for x in queens_adult_care_serv_neigh_list if x not in p]
print("Not in queens_pharmacies_and_parks_neigh_list: ", queens_adult_care_serv_neigh_not_in_ph_and_park_list)

Not in queens_pharmacies_and_parks_neigh_list: ['Bellerose', 'Breezy Point', 'Briarwood', 'East Elmhurst', 'East Flushing', 'Far Rockaway', 'Flushing', 'Ft. Totten', 'Hammets', 'Jamaica', 'Jamaica Estates', 'Murray Hill', 'North Corona', 'Queens Village', 'St. Albans']
```



Distance calculation using Haversine formula and final data frame

For each park we did calculate the distance between the park and all pharmacies in Queens and the park and all existing adult day care centers in Queens distributed across clusters and limited the output to pharmacies within 2 km from the park and to adult care services within 1 km from the park. After this step we created the final data frame that contains for each park respective counts of pharmacies within 2 km from the park and counts of adult day care centers within 1 km from the park along with respective cluster labels.


```
In [111]: queens_final_parks_pharm_adult_serv_info = queens_parks[['Neighborhood', 'Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category', 'Cluster Labels']]
queens_final_parks_pharm_adult_serv_info.rename(columns={'Venue Latitude': 'Park Latitude', 'Venue Longitude': 'Park Longitude'}, inplace=True)
queens_final_parks_pharm_adult_serv_info = pd.merge(queens_final_parks_pharm_adult_serv_info, pacsdk_1km_or_less_stats, how='left', left_on=['Park Latitude', 'Park Longitude'], right_on=['Park Latitude', 'Park Longitude'])
queens_final_parks_pharm_adult_serv_info.rename(columns={'cnt': 'AdultService Count'}, inplace=True)
queens_final_parks_pharm_adult_serv_info = pd.merge(queens_final_parks_pharm_adult_serv_info, pphsd_2km_or_less_stats, how='left', left_on=['Park Latitude', 'Park Longitude'], right_on=['Park Latitude', 'Park Longitude'])
queens_final_parks_pharm_adult_serv_info.rename(columns={'cnt': 'Pharmacy Count'}, inplace=True)
values = {'AdultService Count': 0, 'Pharmacy Count': 0}
queens_final_parks_pharm_adult_serv_info.fillna(value=values, inplace=True)
queens_final_parks_pharm_adult_serv_info = queens_final_parks_pharm_adult_serv_info.astype({'AdultService Count': int, 'Pharmacy Count': int})
queens_final_parks_pharm_adult_serv_info
```

Out[111]:

| | Neighborhood | Venue | Park Latitude | Park Longitude | Venue Category | Cluster Labels | AdultService Count | Pharmacy Count |
|----|----------------------|---|---------------|----------------|----------------|----------------|--------------------|----------------|
| 0 | Elmhurst | Broadway Park | 40.740795 | -73.885517 | Park | 1 | 6 | 5 |
| 1 | Corona | William F. Moore Park (Spaghetti Park) | 40.743666 | -73.855443 | Park | 1 | 3 | 3 |
| 2 | Forest Hills | Yellowstone Park | 40.726251 | -73.847759 | Park | 2 | 0 | 6 |
| 3 | Forest Hills | MacDonald Park | 40.722239 | -73.847141 | Park | 2 | 0 | 5 |
| 4 | Kew Gardens | Eight Oaks Triangle | 40.707226 | -73.827127 | Park | 2 | 1 | 2 |
| 5 | Kew Gardens | LIRR Metropolitan Yard | 40.703840 | -73.824317 | Park | 2 | 1 | 1 |
| 6 | Richmond Hill | Li Frank McConnell Park | 40.694270 | -73.829010 | Park | 1 | 2 | 4 |
| 7 | Sunnyside | Thomas P. Noonan, Jr. Playground | 40.741053 | -73.922213 | Park | 1 | 2 | 5 |
| 8 | Maspeth | Whitefish Triangle Park | 40.726517 | -73.901752 | Park | 2 | 0 | 3 |
| 9 | Rego Park | Fleetwood Triangle | 40.726679 | -73.862636 | Park | 2 | 3 | 5 |
| 10 | Woodhaven | Equity Park | 40.691645 | -73.853378 | Park | 2 | 4 | 5 |
| 11 | South Ozone Park | Back Streets Park (Officer Edward Bym Park) | 40.667846 | -73.806453 | Park | 1 | 0 | 0 |
| 12 | South Ozone Park | Pats Oval Park | 40.668634 | -73.805878 | Park | 1 | 0 | 0 |
| 13 | South Ozone Park | Back Street Park | 40.665542 | -73.806407 | Park | 1 | 0 | 0 |
| 14 | College Point | Popepshausen Park | 40.781653 | -73.844672 | Park | 2 | 0 | 2 |
| 15 | College Point | Popepshausen Triangle Park | 40.788130 | -73.845970 | Park | 2 | 0 | 2 |
| 16 | Glen Oaks | Glen Oaks Oval | 40.748273 | -73.714957 | Park | 2 | 1 | 2 |
| 17 | Hollis | Kings Park | 40.712344 | -73.764489 | Park | 1 | 1 | 0 |
| 18 | Hollis | Jamaica Park | 40.712351 | -73.764478 | Park | 1 | 1 | 0 |
| 19 | Springfield Gardens | Springfield Park | 40.665932 | -73.758064 | Park | 1 | 0 | 2 |
| 20 | Edgemere | Baywater Park | 40.596248 | -73.770970 | Park | 1 | 0 | 0 |
| 21 | Queensboro Hill | Matwee Park | 40.742153 | -73.827568 | Park | 1 | 0 | 0 |
| 22 | Laurelton | Laurelton Park | 40.670598 | -73.735900 | Park | 1 | 0 | 1 |
| 23 | Somerville | community playground | 40.601131 | -73.794403 | Park | 1 | 0 | 0 |
| 24 | Forest Hills Gardens | Hawthorne Park | 40.716422 | -73.840083 | Park | 2 | 0 | 5 |
| 25 | Pomono | Blue park | 40.733952 | -73.808954 | Park | 1 | 5 | 1 |
| 26 | Hunters Point | Hunter's Point Community Park | 40.745033 | -73.953225 | Park | 1 | 0 | 0 |
| 27 | Sunnyside Gardens | Torsney Playground | 40.747019 | -73.921128 | Park | 2 | 2 | 5 |
| 28 | Middle Village | Juniper Valley Park | 40.720281 | -73.881258 | Park | 2 | 0 | 3 |
| 29 | Baywater | Inwood Park | 40.614236 | -73.761475 | Park | 1 | 0 | 0 |
| 30 | Queensbridge | Queensbridge Park | 40.756701 | -73.948653 | Park | 1 | 0 | 0 |

The solution of our problem are parks with Adult Service Count = 0 and Pharmacy Count > 0. There are 9 parks satisfying these 2 conditions:

```
In [112]: queens_final_parks_pharm_adult_serv_candidates = queens_final_parks_pharm_adult_serv_info[(queens_final_parks_pharm_adult_serv_info["AdultService Count"]==0) & (queens_final_parks_pharm_adult_serv_info["Pharmacy Count"] > 0)]
queens_final_parks_pharm_adult_serv_candidates.potential = queens_final_parks_pharm_adult_serv_candidates.sort_values(by=["Cluster Labels", "Pharmacy Count"], ascending=False)
queens_final_parks_pharm_adult_serv_candidates.potential.reset_index(drop=True, inplace=True)
queens_final_parks_pharm_adult_serv_candidates.potential
```

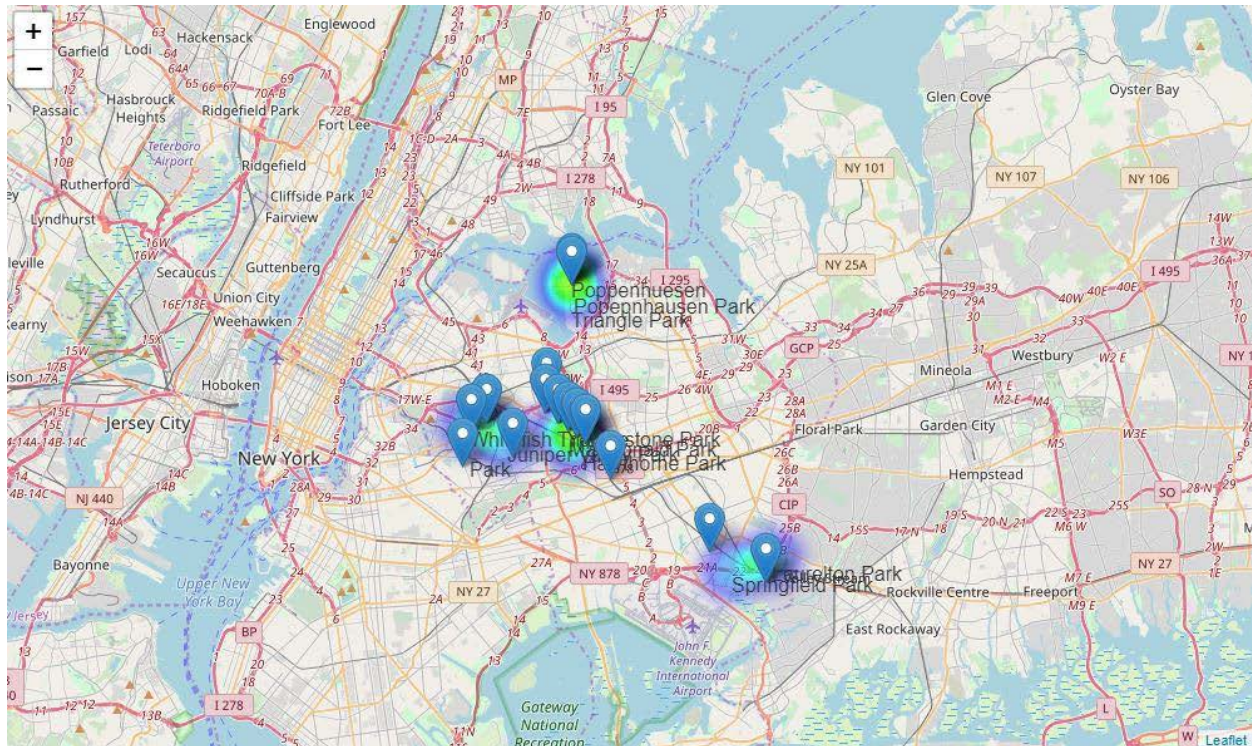
Out[112]:

| | Neighborhood | Venue | Park Latitude | Park Longitude | Venue Category | Cluster Labels | AdultService Count | Pharmacy Count |
|---|----------------------|----------------------------|---------------|----------------|----------------|----------------|--------------------|----------------|
| 0 | Forest Hills | Yellowstone Park | 40.726251 | -73.847759 | Park | 2 | 0 | 6 |
| 1 | Forest Hills | MacDonald Park | 40.722239 | -73.847141 | Park | 2 | 0 | 5 |
| 2 | Forest Hills Gardens | Hawthorne Park | 40.716422 | -73.840083 | Park | 2 | 0 | 5 |
| 3 | Maspeth | Whitefish Triangle Park | 40.726517 | -73.901752 | Park | 2 | 0 | 3 |
| 4 | Middle Village | Juniper Valley Park | 40.720281 | -73.881258 | Park | 2 | 0 | 3 |
| 5 | College Point | Popepshausen Park | 40.781653 | -73.844672 | Park | 2 | 0 | 2 |
| 6 | College Point | Popepshausen Triangle Park | 40.788130 | -73.845970 | Park | 2 | 0 | 2 |
| 7 | Springfield Gardens | Springfield Park | 40.665932 | -73.758064 | Park | 1 | 0 | 2 |
| 8 | Laurelton | Laurelton Park | 40.670598 | -73.735900 | Park | 1 | 0 | 1 |

Out[114]:

| | Neighborhood | Park | Pharmacy Latitude | Pharmacy Longitude | Venue | Distance |
|----|----------------------|----------------------------|-------------------|--------------------|----------------------|-------------|
| 11 | Forest Hills Gardens | Hawthorne Park | 40.718438 | -73.838177 | Rite Aid | 275.817504 |
| 12 | Forest Hills Gardens | Hawthorne Park | 40.721396 | -73.843421 | CVS pharmacy | 620.683159 |
| 13 | Forest Hills Gardens | Hawthorne Park | 40.724004 | -73.847911 | Walgreens | 1070.793864 |
| 14 | Forest Hills Gardens | Hawthorne Park | 40.726791 | -73.853772 | CVS pharmacy | 1631.436981 |
| 15 | Forest Hills Gardens | Hawthorne Park | 40.703557 | -73.824861 | CVS pharmacy | 1922.180490 |
| 19 | Middle Village | Juniper Valley Park | 40.712905 | -73.878023 | AJ's Village Chemist | 864.574704 |
| 20 | Middle Village | Juniper Valley Park | 40.727184 | -73.892861 | CVS pharmacy | 1243.460892 |
| 21 | Middle Village | Juniper Valley Park | 40.722987 | -73.900637 | Eagle pharmacy | 1661.073463 |
| 28 | Laurelton | Laurelton Park | 40.660851 | -73.739247 | Walgreens | 1120.385569 |
| 6 | Forest Hills | MacDonald Park | 40.724004 | -73.847911 | Walgreens | 206.755769 |
| 7 | Forest Hills | MacDonald Park | 40.721396 | -73.843421 | CVS pharmacy | 327.295251 |
| 8 | Forest Hills | MacDonald Park | 40.726791 | -73.853772 | CVS pharmacy | 754.126679 |
| 9 | Forest Hills | MacDonald Park | 40.718438 | -73.838177 | Rite Aid | 865.907530 |
| 10 | Forest Hills | MacDonald Park | 40.730898 | -73.860729 | CVS pharmacy | 1496.432745 |
| 22 | College Point | Popepnhansen Park | 40.783961 | -73.846040 | Rite Aid | 281.357993 |
| 23 | College Point | Popepnhansen Park | 40.785310 | -73.845660 | Walgreens | 415.166141 |
| 24 | College Point | Poppenhuesen Triangle Park | 40.785310 | -73.845660 | Walgreens | 314.674123 |
| 25 | College Point | Poppenhuesen Triangle Park | 40.783961 | -73.846040 | Rite Aid | 463.686857 |
| 26 | Springfield Gardens | Springfield Park | 40.673555 | -73.770720 | Variety Drugs | 1363.404179 |
| 27 | Springfield Gardens | Springfield Park | 40.660851 | -73.739247 | Walgreens | 1685.246699 |
| 16 | Maspeth | Whitefish Triangle Park | 40.722987 | -73.900637 | Eagle pharmacy | 403.731647 |
| 17 | Maspeth | Whitefish Triangle Park | 40.727184 | -73.892861 | CVS pharmacy | 753.037221 |
| 18 | Maspeth | Whitefish Triangle Park | 40.708905 | -73.905848 | Rite Aid | 1989.080718 |
| 0 | Forest Hills | Yellowstone Park | 40.724004 | -73.847911 | Walgreens | 250.195391 |
| 1 | Forest Hills | Yellowstone Park | 40.726791 | -73.853772 | CVS pharmacy | 510.425139 |
| 2 | Forest Hills | Yellowstone Park | 40.721396 | -73.843421 | CVS pharmacy | 652.088138 |
| 3 | Forest Hills | Yellowstone Park | 40.718438 | -73.838177 | Rite Aid | 1186.376035 |
| 4 | Forest Hills | Yellowstone Park | 40.730898 | -73.860729 | CVS pharmacy | 1209.297584 |
| 5 | Forest Hills | Yellowstone Park | 40.737679 | -73.859863 | Rite Aid | 1629.929919 |

Here is the heat map based on the count of respective pharmacies to graphically display promising locations. Popups mark pharmacies locations.



It is important to note here even if we use directly the `queens_adult_care_services_data` without performing a join to find out to which cluster respective adult day care center belongs the solution is still the same. See the notebook for details. In other words our solution is not sensitive to splitting the NTA column on "-".

Conclusion

The purpose of this project was to find an optimal location for an adult day care center in the Queens Borough of New York City whose main focus would be on recreational activities and social stimulation for the elderly population who would otherwise stay at home alone. The optimal location had to satisfy 2 conditions:

1. No existing adult day care center within 1 km of the park latitude and longitude coordinates.
2. At least one pharmacy within 2 km of the park latitude and longitude coordinates.

We did show that it's possible to identify such locations using the Foursquare location data in combination with the Queens neighborhoods and adult day care centers data. 12

The k-means clustering can provide a quick way to identify most of the potential candidate parks. These park locations will be presented to stakeholders as a starting point for stakeholders final 'street level' exploration to determine the optimal adult day care center location. Final decision on optimal adult day care center location will be made by the stakeholders who would also need to take into consideration additional factors such as federal and state requirements, medical and insurance requirements, zoning laws and real estate availability information and prices around the parks. Distance limits of the adult day care centers and pharmacies used in this project for the optimal location are not set in stone and can be customized based on the stakeholders interest. Future directions include researching how to incorporate real estate availability information and prices data around the parks to produce more focused locations.