Capstone Project - The Battle of Neighborhoods (Week 2)

Problem statement

- Healthy food, vegetrian restaurants are of prime importance for preventive healthcare
  - Starting a vegterian restaurant can be a great business opportunity

- By emphesizing on the locations where gym centers are popular will lead us to the right place of our planned vegetarian restaurants
- Provide this study for the city of New York and in particular to Manhattan neighbourhood
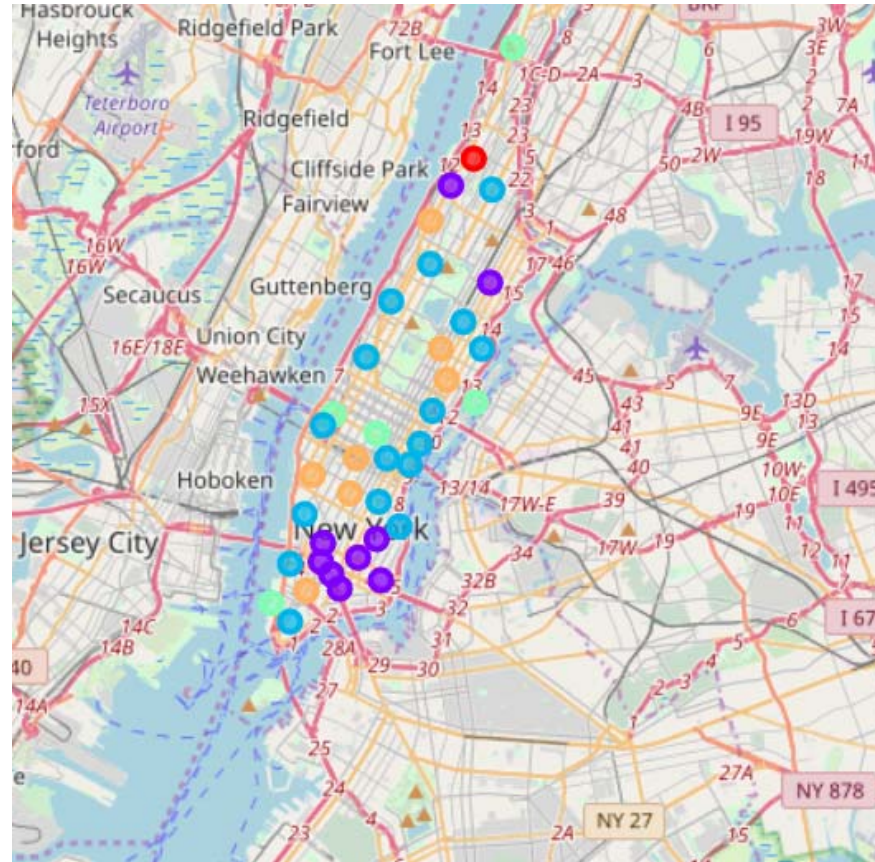- Use Machine learning techniques

Dataset

- Publicly available data of New York city available online: https://geo.nyu.edu/catalog/nyu_2451_34572

- 5 boroughs and 306 neighborhoods

- Use https://developer.foursquare.com/docs/resources/categories in order to find gym/fitness center category id

Methodology

- Loading and cleaning the dataset
- Performing k-means statisical analysis in two stages
  - we consider venues with category = gym/fitness
  - clustering regardless of the category type
- Decide if a neighborhood needs a vegetarian restaurant, by inspecting if
  - it appears in a cluster with high concentation in gym/fitness venues (i.e. by looking for clustering in Stage 1)
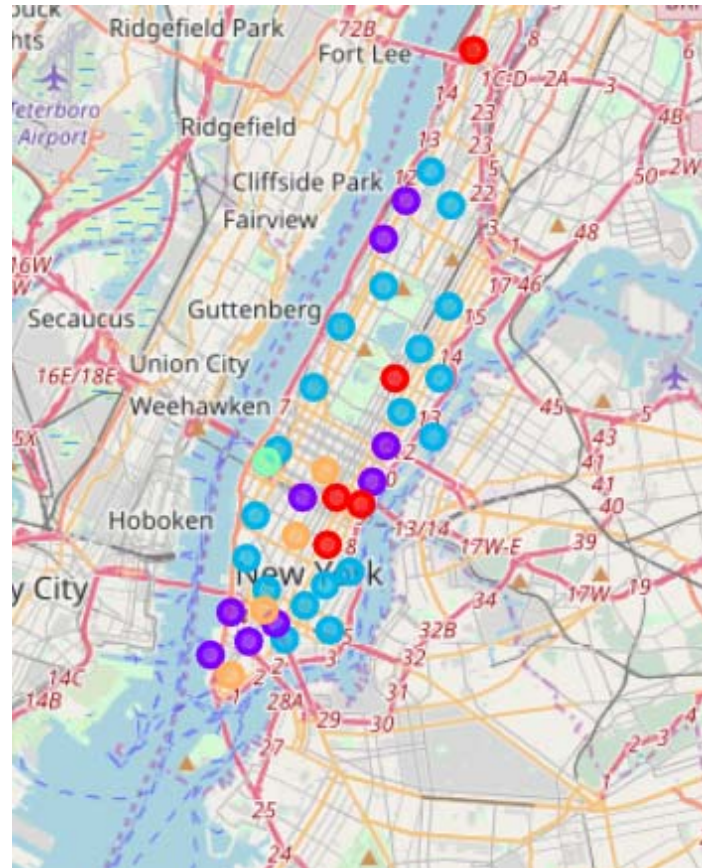 - and low concentration in restaurants (stage 2)

Results

- Clustering/Stage 1

Results

- Clustering/stage 2

Results

- Applying our methodology to the results by taking a use case:
  - 'Chelsea neighborhood'
- In Stage 1: 'Chelsea' belongs to cluster 5 (high concentration in gym)

Results

- Stage 2: Chelsea appears in cluster3 (low concentration in healthy food restaurants) ➔ Chelsea is a good candidate

Cluster 3

```
In [51]: tan_merged_all['Cluster Labels'] == 2, manhattan_merged_all.columns[[1] + list(range(5, manhattan_all...
<
```

Out[51]:

	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
0	Marble Hill	Deli / Bodega	Salon / Barbershop	Residential Building (Apartment / Condo)	High School	Laundry Service	Church	Furniture / Home Store	Nail Salon	Office
1	Chinatown	Chinese Restaurant	Miscellaneous Shop	Bus Station	Bakery	Bridge	Noodle House	Park	Arts & Crafts Store	Bus Line
3	Inwood	Laundry Service	Deli / Bodega	Salon / Barbershop	Nail Salon	American Restaurant	Bank	Mexican Restaurant	Pharmacy	Bus Line
4	Hamilton Heights	Residential Building (Apartment / Condo)	Salon / Barbershop	Building	Deli / Bodega	Non-Profit	Laundry Service	Church	Bar	Bus Station
13	Lincoln Square	Opera House	High School	Theater	Art Gallery	Residential Building (Apartment / Condo)	Building	Performing Arts Venue	Library	Event Space	Office
14	Clinton	Building	Bus Station	Theater	Bus Line	Lounge	Residential Building (Apartment / Condo)	Restaurant	Gym / Fitness Center	Deli / Bodega	Cocktail Bar
17	Chelsea	High School	Building	Residential Building (Apartment / Condo)	Deli / Bodega	Office	Art Gallery	Pet Store	Pharmacy	Music Venue	Laundry Service
18	Greenwich Village	Residential Building (Apartment / Condo)	Salon / Barbershop	Art Gallery	Furniture / Home Store	Cocktail Bar	Jewelry Store	Kids Store	Tech Startup	Ice Cream Shop	Clothing Store

Conclusion

- Machine learning is considered a core leveraging technology
- We applied a very simple unsupervised machine learning algorithm (i.e. Kmeans) in order to solve a business problem
  - opening a new vegetarian restaurant in the most appropriate neighborhood of New York city/ Manhattan borough
- Using one use case: our methodology is successful in finding a perfect location for our planned business.