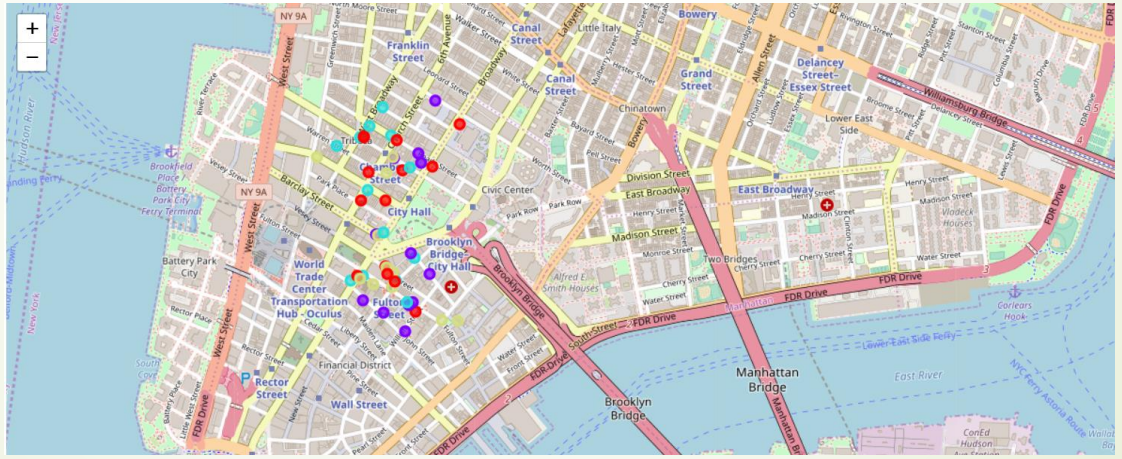
Finding Good Quality Food in New York City

Introduction

- Whenever some one visit or move to a new place, you find it hard to find good place to eat. I have explored New York city to find good quality RESTAURANTS using k-means algorithms.
- I have used foursquare API to find restaurants and their reviews.



Methodology

- The thought process behind this is that likes are a proxy for quality. The more likes there are, the better the restaurant is. I will then bin this data into a quality categorical variables so we can cluster appropriately.

- I am also going to create new categorical variables for the restaurants to better group them based on type of cuisine. This way you can look for good Mexican food or now what type of food might be best to eat in New York if you are new to the area.

- Using geolocator I have found the coordinates of New York cities and than using those coordinates retrieved all the venue information from the Foursquare API

Out[101]:

	venue.name	venue.id	venue.categories	venue.location.lat	venue.location.lng
0	The Bar Room at Temple Court	57f0689d498e7d49d9189369	Hotel Bar	40.711448	-74.006802
1	The Beekman - A Thompson Hotel	56d8c0f8498edb854f926e6a	Hotel	40.711173	-74.006702
2	Alba Dry Cleaner & Tailor	4c606c3e1e5cd13ad1a1a1ed	Laundry Service	40.711434	-74.006272
3	City Hall Park	3fd66200f964a520d8f11ee3	Park	40.712415	-74.006724
4	Gibney Dance Center Downtown	53373f26498e940581c90985	Dance Studio	40.713923	-74.005661
5	Augustine	58191674ded8f8626ed70af0	French Restaurant	40.711310	-74.006660
6	The Wooly Daily	56093809498e5344ab8835a6	Coffee Shop	40.712137	-74.008395
7	The Class by Taryn Toomey	58a253c830ecc66c9e5b40a0	Gym / Fitness Center	40.712753	-74.008734
8	CrossFit 212 TriBeCa	52001eed498e9ac16ca5e20b	Gym	40.714537	-74.005999
9	Takahachi Bakery	4c154c9a77cea593c401d260	Bakery	40.713653	-74.008804
10	Aahar Indian Cuisine	575dea4c498e2739e43a27e2	Indian Restaurant	40.713307	-74.007994
11	The Hummus & Pita Co	515c4ca3e4b00279706b095d	Falafel Restaurant	40.714453	-74.006853

- Now I filtered all the venues , getting the data frame with food restaurants/café/Bakery…

- I have divided all the food restaurants into different cusines so that we are able to identify the type of cuisines that restaurant has to offer.
- I have also binned the data according to likes and categorize them as Below Average, Poor, Average, Great

```
# let's create our new categories and create a function to apply those to our existing data

European_food= ['French Restaurant','Pizza Place','Italian Restaurant']
South_American_food= ['Taco Place','Cuban Restaurant', 'Burrito Place','Deli / Bodega']
South_East_Asian_food=['Indian Restaurant','Vegetarian / Vegan Restaurant']
Middle_East_food=['Falafel Restaurant','Middle Eastern Restaurant','Greek Restaurant']
East_Asian_food=['Japanese Restaurant','Sushi Restaurant','Japanese Curry Restaurant']
American_food=['American Restaurant','Burger Joint','Restaurant','Fast Food Restaurant']
Other=['Coffee Shop','Bakery','Sandwich Place','Café','Juice Bar','Cocktail Bar','Molecular Gastronomy Restaurant','Bubble Tea Shop','Salad Place']
bars=['Bar']
```

- After that I applied One Hot encoding to the Categories and likes column and created a new dataframe and applied k-means algorithm.

Results

- Running my clustering algorithm, I got four clusters of restaurants. These are as follows:
  - Cluster #1: Poor Quality Restaurants
  - Cluster #2: Average Quality Restaurants
  - Cluster #3: Great Quality Restaurants
  - Cluster #4: Below Average Quality Restaurants

Conclusion

- There are a variety of restaurants available in New York City. I tried to cluster them according to their likings, but if we increase the cluster size we might able to get rankings according to different cuisines which will be more helpful to our target audience.