Exploring the Taste of NYC Neighborhoods

1a. Introduction

New York City is the most populous city in the United States, home to the headquarters of the United Nations and an important center for international diplomacy. It just might be the most diverse city on the planet, as it is home to over 8.6 million people and over 800 languages.

"Traditional cuisine is passed down from one generation to the next. It also operates as an expression of cultural identity. Immigrants bring the food of their countries with them wherever they go and cooking traditional food is a way of preserving their culture when they move to new places."

Undoubtedly, **Food Diversity** is an important part of an ethnically diverse metropolis. The idea of this project is to categorically segment the neighborhoods of New York City into major clusters and examine their cuisines. A desirable intention is to examine the neighborhood cluster's food habits and taste. Further examination might reveal if food has any relationship with the diversity of a neighborhood.

This project will help to understand the diversity of a neighborhood by leveraging venue data from Foursquare's 'Places API' and 'k-means clustering' machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighborhood.
Stakeholders would be the one who are interested to use this quantifiable analysis to understand the distribution of different cultures and cuisines over "the most diverse city on the planet - NYC". Also, this project can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.

1b. Business Problem

New York City and the city of Toronto. are very diverse and are the financial capitals of their respective countries. One interesting idea would be to compare the neighborhoods of the two cities and determine how similar or dissimilar they are. Is New York City more like Toronto or Paris or some other multicultural city? if someone is looking to open a restaurant, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

2. Data

To examine the above said, following data sources will be used:

1. **New York City Dataset**
    - Description: This New York City Neighborhood Names point file was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale. This dataset will provide the addresses of neighborhood of NYC in json format. An extract of the json is as follows:
    - {'type': 'Feature',
    - 'id': 'nyu_2451_34572.306',
    - 'geometry': {'type': 'Point',
    - 'coordinates': [-74.08173992211962, 40.61731079252983]},
    - 'geometry_name': 'geom',
    - 'properties': {'name': 'Fox Hills',
    - 'stacked': 2,
    - 'annoline1': 'Fox',
    - 'annoline2': 'Hills',
    - 'annoline3': None,
    - 'annoangle': 0.0,
    - 'borough': 'Staten Island',
    - 'bbox': [-74.08173992211962,

- 40.61731079252983,
- -74.08173992211962,
  40.61731079252983]}}

2. **Foursquare API:**
   - Description: Foursquare API, a location data provider, will be used to make RESTful API calls to retrieve data about venues in different neighborhoods. This is the link to Foursquare Venue Category Hierarchy. Venues retrieved from all the neighborhoods are categorized broadly into "Arts & Entertainment", "College & University", "Event", "Food", "Nightlife Spot", "Outdoors & Recreation", etc. An extract of an API call is as follows:
   - 'categories': [{'id': '4bf58dd8d48988d110941735',
   - 'name': 'Italian Restaurant',
   - 'pluralName': 'Italian Restaurants',
   - 'shortName': 'Italian',
   - 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/italian_',
   - 'suffix': '.png'},
   - 'primary': True}],
   - 'verified': False,
   - 'stats': {'tipCount': 17},
   - 'url': 'http://eccorestaurantny.com',
     'price': {'tier': 4, 'message': 'Very Expensive', 'currency'

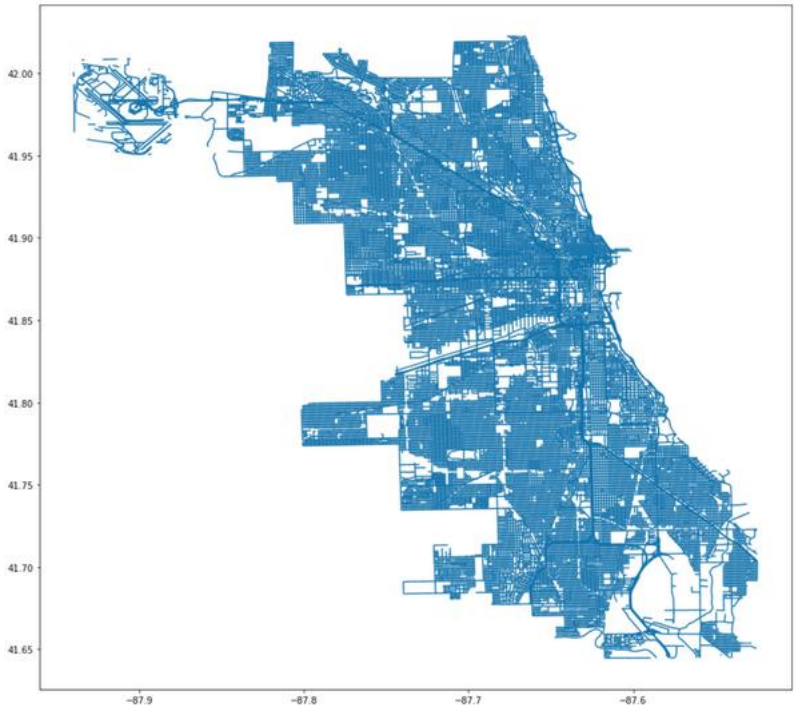
2. Methodology

Download and Explore New York City Dataset

In order to segment the neighborhoods of New York City, a dataset is required that contains the 5 boroughs and the neighborhoods, that exist in each borough, with respective latitude and longitude coordinates.

Dataframe:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585



3. Four Square Credentials and Version:

Foursquare API provides a range of tools for developers to incorporate the up-to-date location data to enhance their projects.

To start using Foursquare, first, let's create a new account at https://developer.foursquare.com/ . The default account is called *Sandbox Account*. Sandbox Account provides you a limited set of tools, for instance, you'll be able to make only 950 Regular Calls / Day, and get 1 Photo and 1 Tip per Venue. Not too much, but it should be enough for a "toy" dataset (sure, you may also upgrade your account for future projects). Foursquare provides a simple UI to manage your account and to check your daily stats (like how many calls did you make). Once the account is created, you will be assigned credentials: the *Client ID and the Client Secret.*

Ok, all set up and we can jump into coding. First and foremost, let's import some necessary libraries: ***pandas.io.json(***to tranforming json file into a pandas dataframe library), ***folium*** (plotting library), ***geopy.geocoders*** (module to convert an address into latitude and longitude values), and ***requests*** (library to handle requests).

4. *Food* & It's Sub-Categories:

Let's create a function to return a dictionary with Ids' & Names of *Food* & It's Sub-Categories.

The category name of the venue **'Carvel Ice Cream'** is **'Food'**.

As, our aim is to segment the neighborhoods of NYC with respect to the *Food* in its vicinity. We need to proceed further to fetch this data from all the 306 neighborhoods' venues.

Let's create a function to repeat the following process to all the neighborhoods in NYC:

- Loop through neighborhoods
    - Create the API request URL with radius=500, LIMIT=100
    - Make the GET request
    - For each neighborhood, return only relevant information for each nearby venue
    - Append all nearby venues to a list
- Unfold the list & append it to dataframe being returned

5. **Exploratory Analysis:**

Pandas is a Python library that provides extensive means for data analysis. Data scientists often work with data stored in table formats like .csv, .tsv, or .xlsx. Pandas makes it very convenient to load, process, and analyze such tabular data using SQL-like queries. In conjunction with Matplotlib and Seaborn, Pandas provides a wide range of opportunities for visual analysis of tabular data.

The main data structures in Pandas are implemented with **Series** and **DataFrame** classes. The former is a one-dimensional indexed array of some fixed data type. The latter is a two-dimensional data structure - a table - where each column contains data of the same type. You can see it as a dictionary of Series instances. DataFrames are great for representing real data: rows correspond to instances (examples, observations, etc.), and columns correspond to features of these instances.

- There are 194 uniques categories.
- Out[29]:
- Venue Category
- Deli / Bodega 1136
- Pizza Place 1078
- Coffee Shop 919
- Donut Shop 710
- Fast Food Restaurant 664
- Chinese Restaurant 619
- Italian Restaurant 544
- Bakery 544
- American Restaurant 428
- Café 401
- Caribbean Restaurant 359
- Bagel Shop 357
- Mexican Restaurant 352
- Sandwich Place 332
- Diner 313
- Ice Cream Shop 256
- Fried Chicken Joint 254
- Restaurant 224
- Food 204
- Burger Joint 204
- Seafood Restaurant 155
- Sushi Restaurant 152
- Latin American Restaurant 144
- Asian Restaurant 142
- Spanish Restaurant 140
- Japanese Restaurant 136
- Food Truck 136
- Bar 120
- Juice Bar 111
- New American Restaurant 108
- Dessert Shop 104
- Breakfast Spot 99
- Indian Restaurant 98
- Food Court 95
- Thai Restaurant 94
- Taco Place 92
- Korean Restaurant 88
- BBQ Joint 88
- Salad Place 75

- French Restaurant 72
- Bubble Tea Shop 72
- Grocery Store 70
- Middle Eastern Restaurant 68
- Mediterranean Restaurant 63
- Greek Restaurant 60
- Steakhouse 59
- Vegetarian / Vegan Restaurant 57
- Cocktail Bar 56
- Southern / Soul Food Restaurant 55
- Vietnamese Restaurant 54
- Gastropub 53
- Ramen Restaurant 46
- Wings Joint 42
- South American Restaurant 34
- Peruvian Restaurant 30
- Filipino Restaurant 29
- Halal Restaurant 28
- Gourmet Shop 27
- Snack Place 27
- Turkish Restaurant 27
- Tapas Restaurant 25
- Frozen Yogurt Shop 25
- African Restaurant 24
- Beer Garden 24
- Hot Dog Joint 24
- Pub 23
- Cantonese Restaurant 23
- Comfort Food Restaurant 22
- Eastern European Restaurant 20
- Dumpling Restaurant 20
- Fish & Chips Shop 18
- German Restaurant 18
- Hotel 18
- Dim Sum Restaurant 17
- Convenience Store 17
- Lounge 17
- Cuban Restaurant 17
- Noodle House 16
- Burrito Place 15
- Tea Room 15
- Sports Bar 14
- Empanada Restaurant 14
- Wine Bar 14
- Falafel Restaurant 14
- Bistro 13
- Cafeteria 13
- Sri Lankan Restaurant 13
- Shanghai Restaurant 13
- Russian Restaurant 12
- Irish Pub 12
- Australian Restaurant 12
- Food & Drink Shop 11
- Hotpot Restaurant 10
- Udon Restaurant 10
- Supermarket 10
- Bookstore 10

- Buffet 10
- Gym / Fitness Center 9
- Arepa Restaurant 9
- Theme Restaurant 9
- Cajun / Creole Restaurant 9
- Farmers Market 8
- Health Food Store 8
- Malay Restaurant 8
- Tibetan Restaurant 8
- Cupcake Shop 8
- Flea Market 7
- Taiwanese Restaurant 7
- Colombian Restaurant 7
- Moroccan Restaurant 7
- Poke Place 6
- Event Space 6
- Art Gallery 6
- Club House 5
- Arcade 5
- Himalayan Restaurant 5
- Smoothie Shop 5
- Hotel Bar 5
- Ukrainian Restaurant 5
- Gas Station 5
- Gift Shop 5
- Tex-Mex Restaurant 5
- Pakistani Restaurant 5
- Israeli Restaurant 5
- Other Nightlife 5
- Mac & Cheese Joint 5
- Hookah Bar 5
- Kosher Restaurant 4
- Beer Bar 4
- Soup Place 4
- Residential Building (Apartment / Condo) 4
- Brazilian Restaurant 4
- Music Venue 4
- Pet Café 4
- Polish Restaurant 4
- Gluten-free Restaurant 4
- Argentinian Restaurant 4
- Persian Restaurant 4
- Austrian Restaurant 4
- Pie Shop 3
- Financial or Legal Service 3
- College Cafeteria 3
- Fish Market 3
- Modern European Restaurant 3
- Creperie 3
- Salvadoran Restaurant 3
- Lebanese Restaurant 3
- South Indian Restaurant 3
- Indian Chinese Restaurant 3
- Food Stand 3
- Ethiopian Restaurant 3
- Cheese Shop 3
- Organic Grocery 2

- Shabu-Shabu Restaurant 2
- Rock Club 2
- Beach Bar 2
- Skating Rink 2
- Street Food Gathering 2
- Szechuan Restaurant 2
- Butcher 2
- Afghan Restaurant 2
- Candy Store 2
- Modern Greek Restaurant 2
- Government Building 2
- English Restaurant 2
- Indie Movie Theater 2
- Jewish Restaurant 2
- Market 2
- Miscellaneous Shop 2
- Laundry Service 2
- Paella Restaurant 2
- Nightclub 1
- Portuguese Restaurant 1
- Venezuelan Restaurant 1
- Varenyky restaurant 1
- Furniture / Home Store 1
- Flower Shop 1
- Pastry Shop 1
- Hawaiian Restaurant 1
- Hong Kong Restaurant 1
- Factory 1
- Caucasian Restaurant 1
- Burmese Restaurant 1
- Dosa Place 1
- Kebab Restaurant 1
- Beach 1
- Dive Bar 1
- Park 1
- Clothing Store 1
- Building 1
- North Indian Restaurant 1
- Bowling Alley 1
- Moving Target 1
- Fruit & Vegetable Store 1
- Name: Venue Category, dtype: int64

Top 10 Restaurant categories in NYC:

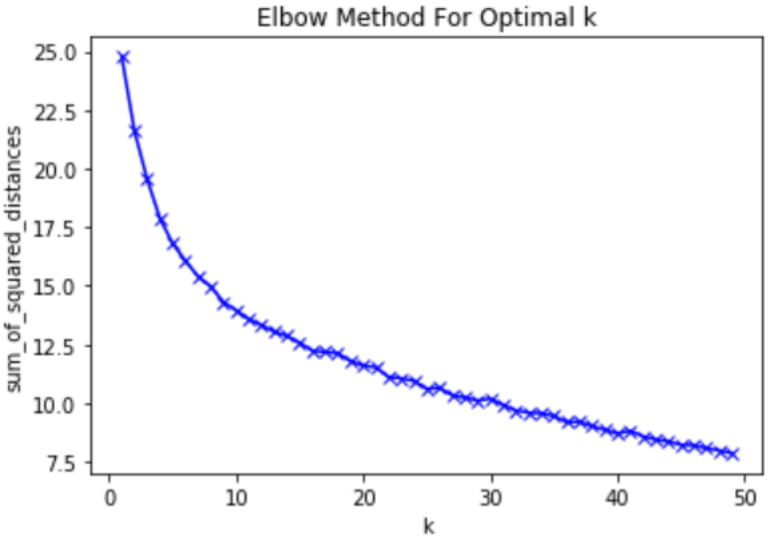
	count	mean	std	min	25%	50%	75%	max
Korean Restaurant	302.0	0.291391	1.828491	0.0	0.0	0.0	0.0	29.0
Chinese Restaurant	302.0	2.049669	2.083199	0.0	1.0	2.0	3.0	17.0
Caribbean Restaurant	302.0	1.188742	2.785965	0.0	0.0	0.0	1.0	16.0
Indian Restaurant	302.0	0.324503	1.123885	0.0	0.0	0.0	0.0	15.0
Fast Food Restaurant	302.0	2.198675	2.054150	0.0	1.0	2.0	3.0	11.0
Italian Restaurant	302.0	1.801325	1.983386	0.0	0.0	1.0	3.0	11.0
Pizza Place	302.0	3.569536	2.190314	0.0	2.0	3.0	5.0	10.0
Seafood Restaurant	302.0	0.513245	0.849950	0.0	0.0	0.0	1.0	7.0
New American Restaurant	302.0	0.357616	0.745702	0.0	0.0	0.0	0.0	6.0
Thai Restaurant	302.0	0.311258	0.726210	0.0	0.0	0.0	0.0	6.0

6. **Elbow Method – K Means:**

A fundamental step for any unsupervised algorithm is to determine the optimal number of clusters into which the data may be clustered. The **Elbow Method** is one of the most popular methods to determine this optimal value of k.

We now demonstrate the given method using the K-Means clustering technique using the **Sklearn** library of python.

1. **Distortion:** It is calculated as the average of the squared distances from the cluster centers of the respective clusters. Typically, the Euclidean distance metric is used.
2. **Inertia:** It is the sum of squared distances of samples to their closest cluster center.

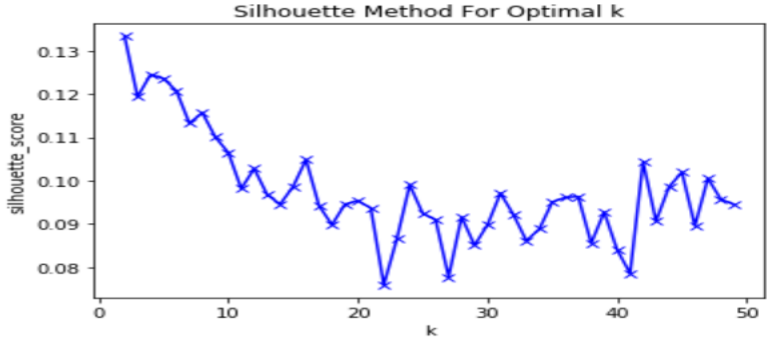


7. **Silhouette Clustering:**

Silhouette refers to a method of interpretation and validation of consistency within clusters of data. The technique provides a succinct graphical representation of how well each object has been classified.[1]

The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters.

The silhouette can be calculated with any distance metric, such as the Euclidean distance or the Manhattan distance.



	Cluster Labels	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	6	Allerton	Mexican Restaurant	Fried Chicken Joint	Pizza Place	Chinese Restaurant	Fast Food Restaurant
1	4	Annadale	Pizza Place	American Restaurant	Sushi Restaurant	Italian Restaurant	Japanese Restaurant
2	4	Arden Heights	Pizza Place	American Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant
3	6	Arlington	Pizza Place	Fast Food Restaurant	American Restaurant	Peruvian Restaurant	Spanish Restaurant
4	2	Arrochar	Italian Restaurant	Pizza Place	Steakhouse	Middle Eastern Restaurant	Chinese Restaurant

8. Conclusion:

Before I end my report, I would like to recap about the Four Square API. using the Foursquare API, we can search for specific type of venues or stores around a given location. And it is important to remember that for this data, we make a regular call to the API, and if you have a free personal developer account ,you can make up to approximately 99 thousand regular calls per day. We can also learn more about a specific venue or store or shop, like their full address, their working hours, and their menu if they have one, and so on. It's also important to remember that for this data, we would need to make a premium call and with the personal developer account, you can make approximately 500 calls per day. Also with the Foursquare API, we can learn more about a specific Foursquare user, their full name, and any tips or photos that they have posted about venues and stores. For this data, a regular call to the API would be made. Furthermore, we can explore a given location by finding what popular spots exist in the vicinity of the location, and for this data a regular call to the API would be made. And finally, with the Foursquare API, we can explore trending venues around a given location.