**1. 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. 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.

**2. Data**

New York City Dataset Link: https://geo.nyu.edu/catalog/nyu\_2451\_34572 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.Foursquare API: Link: https://developer.foursquare.com/docs 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

**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. This dataset exists for free on the web. Here is the link To the dataset: https://geo.nyu.edu/catalog/nyu\_2451\_34572, and To its downloadable json format file: <https://cocl.us/new_york_dataset/newyork_data.json>

!pip install wget

**import** **wget**

**import** **os**

**if** os.path.exists('newyork\_data.json'):

os.remove('newyork\_data.json')

wget.download('https://cocl.us/new\_york\_dataset/newyork\_data.json')

print('**\n**Data downloaded!')

**for** data **in** neighborhoods\_data:

borough = data['properties']['borough']

neighborhood\_name = data['properties']['name']

neighborhood\_latlon = data['geometry']['coordinates']

neighborhood\_lat = neighborhood\_latlon[1]

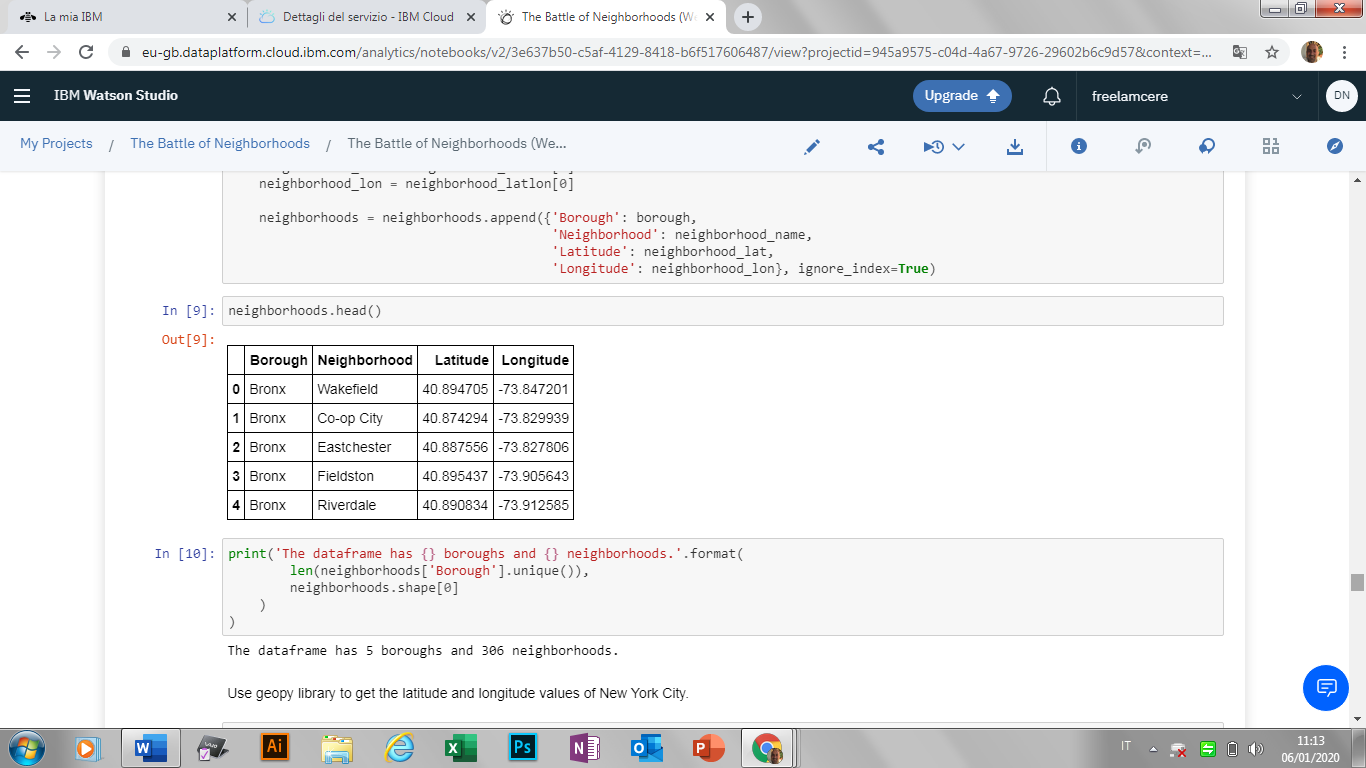
neighborhood\_lon = neighborhood\_latlon[0]

neighborhoods = neighborhoods.append({'Borough': borough,

'Neighborhood': neighborhood\_name,

'Latitude': neighborhood\_lat,

'Longitude': neighborhood\_lon}, ignore\_index=**True**)



*# create map of New York using latitude and longitude values*

map\_newyork = folium.Map(location=[latitude, longitude], zoom\_start=10)

*# add markers to map*

**for** lat, lng, borough, neighborhood **in** zip(neighborhoods['Latitude'], neighborhoods['Longitude'], neighborhoods['Borough'], neighborhoods['Neighborhood']):

label = '**{}**, **{}**'.format(neighborhood, borough)

label = folium.Popup(label, parse\_html=**True**)

folium.CircleMarker(

[lat, lng],

radius=5,

popup=label,

color='blue',

fill=**True**,

fill\_color='#3186cc',

fill\_opacity=0.7,

parse\_html=**False**).add\_to(map\_newyork)

map\_newyork

