New York City Restaurants Cuisine Analysis

AIM:

The goal of the project is to analyze the cuisines of popular restaurants in New York City's neighborhoods. The project shall provide useful insights of New York's restaurant preferences city wide and also specific to each neighborhood of the city using bar graphs and city maps.

The results of this project will reveal the top ten favorite cuisines of the city as well as an insight into city's sub region's cuisine preferences.

The project also aims to cluster the neighborhoods by the cuisine of their popular restaurants using kmeans clustering algorithm to divide and classify the neighborhoods accurately.

The project makes use of foursquare API from the foursquare city guide platform to fetch data of the top 20 restaurants of each neighborhood of the city and the cuisine type of those restaurants. An additional dataset is also required to get the list of all neighborhoods of New York and also their geological coordinates.

Target Audience:

The project will be useful for people who wish to enter the New York restaurant business to understand the present scenario of this business field and help make the best choices such as the cuisine of the restaurant and the location of the restaurant.

The execution of this project is done in Python language using Jupyter notebook. The Matplotlib package is used for data visualization, folium package for plotting maps and SKlearn package for kmeans clustering.

Datasets used for New York restaurants cuisine analysis project

1] Foursquare api from foursquare city guide platform

2] New York City neighborhoods data

3] Geopy package used to convert address to coordinates.

1] Foursquare api from foursquare city guide platform:

We use the foursquare api to acquire the top 20 restaurants in each neighborhood of new York and their cuisine types.

The syntax for fetching for one neighborhood is:

url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, lat, lon, VERSION, search_query, radius, LIMIT)

results = requests.get(url).json()

This returns a json object with detailed information of each of the top restaurants. But this project just needs the name and cuisine from this json object.

The name is in results.response.venues[].name

The cuisine is in results.response.venues[].categories[0].shortname

2] New York City neighborhoods data

We download the dataset from https://geo.nyu.edu/catalog/nyu_2451_34572 or https://cocl.us/new_york_dataset to fetch the names, borough, coordinates of every New York neighborhood.

The dataset looks like this:

	name	borough	lat	lon
0	Wakefield	Bronx	40.8947	-73.8472
1	Co-op City	Bronx	40.8743	-73.8299
2	Eastchester	Bronx	40.8876	-73.8278
3	Fieldston	Bronx	40.8954	-73.9056
4	Riverdale	Bronx	40.8908	-73.9126
...	...	...	...	...
301	Hudson Yards	Manhattan	40.7567	-74.0001
302	Hammels	Queens	40.5873	-73.8055
303	Bayswater	Queens	40.6113	-73.766
304	Queensbridge	Queens	40.7561	-73.9456
305	Fox Hills	Staten Island	40.6173	-74.0817

306 rows × 4 columns

We use the dataset acquired from foursquare api and the above to form the following dataset:

	neighborhood	borough	lat	lon	name	type
0	Wakefield	Bronx	40.894705	-73.847201	Bay 241 Restaurant & Lounge	Caribbean
1	Wakefield	Bronx	40.894705	-73.847201	Kaieteur Restaurant & Bakery	Caribbean
2	Wakefield	Bronx	40.894705	-73.847201	Big Daddy's Caribbean Taste Restaurant	Caribbean
3	Wakefield	Bronx	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	Caribbean
4	Wakefield	Bronx	40.894705	-73.847201	Cool Running Restaurant	Food
5	Wakefield	Bronx	40.894705	-73.847201	Fort Negril Jamaican Restaurant	Food
6	Wakefield	Bronx	40.894705	-73.847201	Allure Restaurant And Bar	Restaurant
7	Wakefield	Bronx	40.894705	-73.847201	Golden Krust Caribbean Restaurant	Caribbean
8	Wakefield	Bronx	40.894705	-73.847201	241 St Cafe & Restaurant	American
9	Wakefield	Bronx	40.894705	-73.847201	Silver Chimes Restaurant	Food

The dataset will further be compressed by grouping rows by neighborhood and one hot encoding it to get no of restaurants in each neighborhood by their cuisines and train the kmeans model to cluster the neighborhoods.

3] Geopy package

The geopy package is used to convert an address to its relevant coordinated we can use in folium map plotting or in foursquare api.

The syntax to convert address "New York" to its coordinates is:

```python
geolocator = Nominatim(user_agent="NY")
location = geolocator.geocode("New York")
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of New York are 40.7127281, -74.0060152.
```

These three are what will be used to get all the data needed for the project.

Methodology:

In the project we do analysis of the cuisines of restaurants that are popular in New York's neighborhoods.

First we find and list of the different cuisines of the restaurants which are popular among the neighborhoods

Then we find the number of popular restaurants of different cuisines in the city. Now that a frequency table has been acquired, the bar graph is plotted for the top ten cuisines. We can now visualize which cuisines are popular in the city by the no of restaurants of that cuisine.

The data is grouped by neighborhood and onehot encoded with the "type" feature (cuisine) . This is now fed into the kmeans clustering model from sklearn package. The number of clusters was chosen as 6 based on the elbow test score graph to determine the most suitable number of cluster. We then try to understand what each cluster is based on their distinguishing features. A map is plotted with circular dots for each neighbor but with colours based on their cluster type or cuisine preference using folium package in python.

Finally bar graphs are plotted for number of neighborhoods the cuisines are most popular .Similar bar graphs are plotted for 2nd most and 3rd most popular graphs.

Results:

Chinese cuisine is highly rated by New Yorkers

popular restaurants of New York's neighborhoods by cuisine
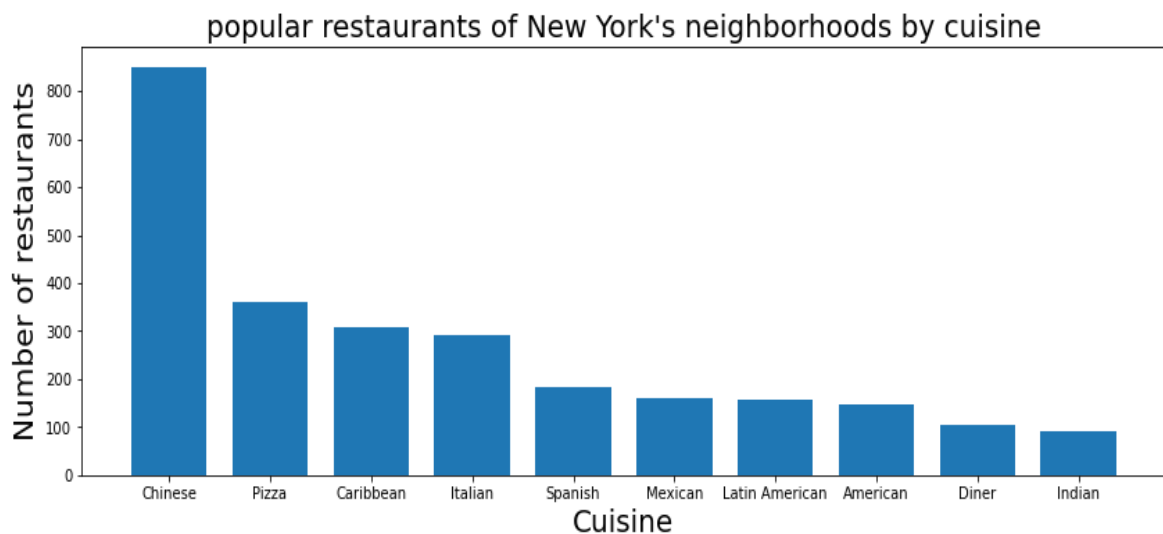
It is quite clear from this graph that New Yorkers rate Chinese cuisines much better than other cuisines. The second place goes for pizza restaurants but is closely contended by Caribbean and Italian cuisine.

Note: The graph must not be misinterpreted that Chinese restaurants are more in numbers. The graph only says that among the popular restaurants of each neighborhood the Chinese cuisine restaurants are more.

The city can be divided geographically based on their preferences:
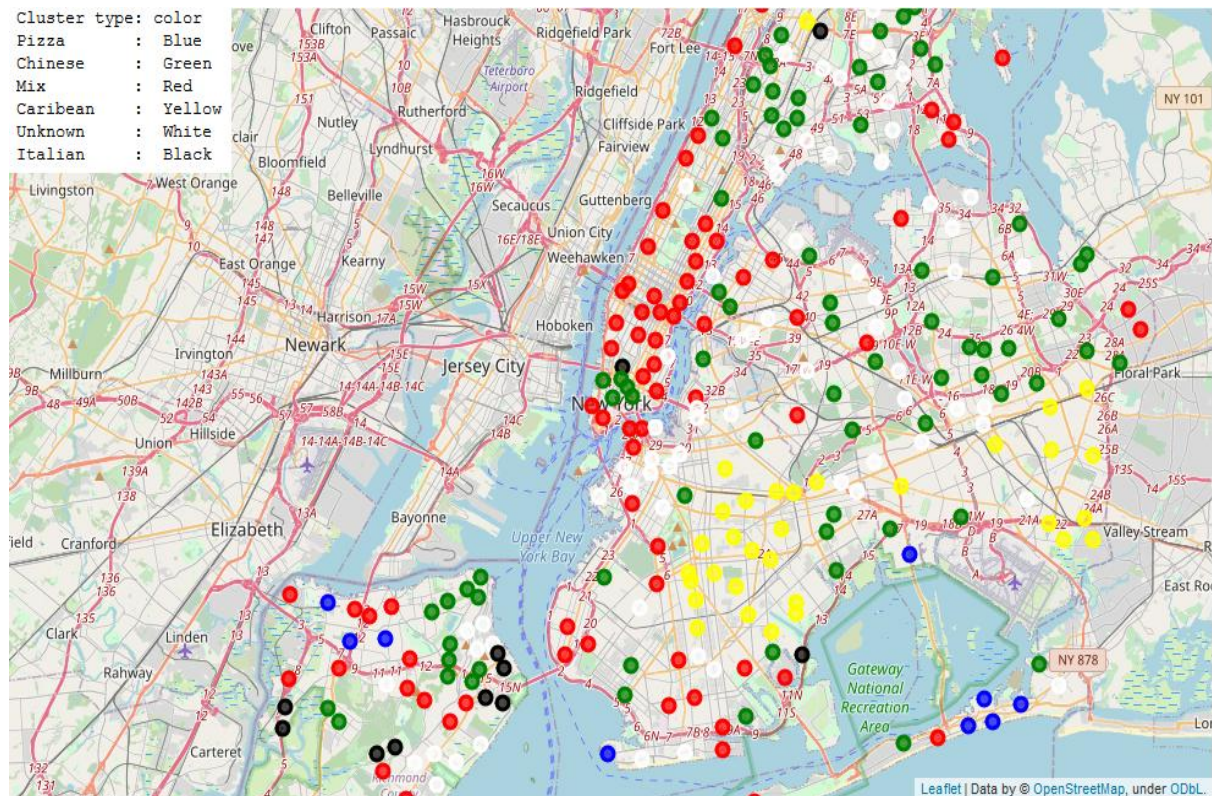
The map shows us that the wide preferences of the city are not evenly distributed.

The Chinese cuisine is very popular in the Bronx borough, the northern half of Queens borough and north west of Staten island.

The Caribbean cuisine is popular in Brooklyn and southern half of Queens and far north of bronx

The Manhattan Island has not shown singular preferences of any cuisine.

Staten Island has shown preference for Italian restaurants with the exception of the northern part of the island.

Discussion:

The Chinese cuisine is undoubtedly the highest rated cuisine in New York however regional preferences cannot be ignored and these are the places where Caribbean, Italian or many different cuisines are preferred.

The foursquare api has a considerable number of restaurants whose cuisines are unknown, They cast doubts on which cuisine among Pizza, Caribbean or Italian holds 2nd place as the race is quite close but numbers are not large enough to dispute that Chinese is highest rated cuisine.

It is recommended that the preferences of the borough or the local region be considered for those who wish to enter the restaurant business in New York.

Conclusion:

City wide the Chinese cuisine is the most highly rated but the city also shows strong local preferences like Caribbean cuisine in Brooklyn, Italian in Staten Island or that Manhattan shows no preference to a single cuisine.