# Capstone project - Caribbean restaurants in New York

### Author: Matej Kuchar

## Introduction

Caribbean cuisine is a fusion of African, Creole, Cajun, Amerindian, European, Latin American, Indian/South Asian, Middle Eastern, and Chinese. These traditions were brought from many different countries when they came to the Caribbean. In addition, the population has created styles that are unique to the region. (Source: [https://en.wikipedia.org/wiki/Caribbean\_cuisine](https://en.wikipedia.org/wiki/Caribbean_cuisine" \t "_blank))

## Business Goal Description

The client wants to expand his network of caribbean restaurants to New York City and is asking to analyze and recommend a few neighborhoods in NY with low density of caribbean restaurants to open a restaurant in.

## Data

For information about the venues, Foursquare API, category "Caribbean Restaurant" (ID 4bf58dd8d48988d144941735) will be used. [https://developer.foursquare.com/docs/build-with-foursquare/categories/](https://developer.foursquare.com/docs/build-with-foursquare/categories/" \t "_blank)

Dataset which contains the New York borroughs and neighborhoods with their latitudes and longitudes will be used. The dataset can be found in the NYU Spatial Data Repository pages: [https://geo.nyu.edu/catalog/nyu\_2451\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572" \t "_blank)

## Methodology

Dataset containing New York borroughs and neighboorhoods with their latitudes and longitudes was used and converted into Pandas dataframe. After that using Foursquare API and Folium map, New York neighboorhoods were explored for their density of caribbean restaurants. Finally using k-means clustering the neighborhoods in New York were divided to 5 clusters based on the density of caribbean restaurants.

## Data exploration and analysis

Using the dataset mentioned in Data section, I transformed the data into Pandas dataframe:

A screenshot of a cell phone

Description automatically generated

Let's get the geographical coorodinates of New York.

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

And now visualize map of New York and its neighboorhoods with Folium library.

A close up of a map

Description automatically generated

Now let's get the data of New York's caribbean restaurants:

A screenshot of a cell phone

Description automatically generated

And visualize the restaurant locations:

A close up of a map

Description automatically generated

Let's compare number of neighborhoods in New York and how many of them have a caribbean restaurant.

There are 306 neighborhoods in New York, yet only in 207 there is a caribbean restaurant.

We will cluster the restaurants in to 5 clusters and then add the cluster labels to our dataset.

A screenshot of a cell phone

Description automatically generated

Let’s explore how many restaurants are in each cluster:

A close up of a logo

Description automatically generated

And visualize the clusters with Folium map:

A close up of a map

Description automatically generated

## Results

Using K-means clustering, New York neighborhoods were divided to 5 clusters based on density of caribbean restaurants and then displayed on the Folium map. Two clusters had really low number of restaurants. Also, in the beginning of the analysis I found out, that there are around 100 neighboorhoods where there is no caribbean restaurant at all. So, to answer the initial request we could recommend places to open the restaurant in.