CAPSTONE PROJECT

Opening Italian Cuisine Business in Brooklyn

OCTOBER 26, 2020

AFRIONI ROMA RIO

Introduction

We have a request from a stakeholder to analyze a potential location in Brooklyn for an Italian cuisine business. Italian cuisines like pizza, pasta, and other cuisines are very famous worldwide and might have potential in a crowded place like Brooklyn, New York City (2.6 million residents). We try to get the result by using unsupervised machine learning k-means and obtain the best position.

Business Problem

This project aims to satisfy the stakeholder's demand, which is to find the best location for an Italian cuisine business in Brooklyn. This project will explain how to obtain the result by using data science methodology, such as unsupervised machine learning k-means.

Data Description

The raw data for analyzing is from this link:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json.

By using a little bit of data preprocessing, we will obtain data set for New York location (borough and neighborhood). After that, we will select borough = Brooklyn for further analysis (see table 1).

Table 1. Example of data set for Brooklyn

Borough	Neighborhood	Latitude	Longitude
Brooklyn	Bay Ridge	40.625801	-74.030621
Brooklyn	Bensonhurst	40.611009	-73.99518
Brooklyn	Sunset Park	40.645103	-74.010316
Brooklyn	Greenpoint	40.730201	-73.954241
Brooklyn	Gravesend	40.59526	-73.973471

To obtain all the venues in Brooklyn, we use foursquare. Then, by using Jupyter Notebook we can obtain the best venues for each neighborhood in Brooklyn. Then, we cluster the cleaned data using a package from scikit-learn