

# **IBM Data Science Capstone Project – The Battle of Neighborhoods**

**Opening an Italian restaurant in Brooklyn, New York  
(Finding optimal neighborhood in Brooklyn for new restaurant)**

By: Zdravko Radulovic

December 2019

Introduction

Business problem

Target Audience

**Data**

## Introduction

Italian restaurants are one of the most popular restaurants in USA. The Italian immigrants who traveled to America in late 19<sup>th</sup>, early 20<sup>th</sup> century bring their culinary culture onto US soil. During World War I Italian immigration paused, and resumed afterward resulting in the establishment of rustic Italian restaurants that catered mainly to immigrants<sup>1</sup>. According to the National Restaurant Association, Italian-American food and Mediterranean cuisine has been highly influential in United States, and it is one of the top three cuisines.<sup>2</sup> Taking this in consideration, opening an Italian restaurant would be great idea. Narrowing this project with Brooklyn borough, and analyzing each neighborhood using Foursquare as a source of data we found that there are 28 Italian restaurants dispersed in neighborhoods of Brooklyn. Finding the right neighborhoods for potential location of a new Italian restaurant would be a key factor that will determine whether the restaurant will be successful investment opportunity or not.

## Business problem

The objective of this project is to analyze neighborhoods in Brooklyn and propose possible location for opening new Italian restaurant. Using data science techniques, we will try to answer the business question: *What are the best neighborhoods for potential opening a new Italian restaurant?*

## Target Audience

Choosing location is key to success for almost any business, especially restaurant business. So this project is useful for potential Italian restaurant chains or restaurant investors that are looking to open or invest in new restaurant in neighborhood that lacks this type of restaurants.

## Data

For this project we need following data:

1. List of all neighborhoods in Brooklyn
2. Venue data, related to Italian restaurants located in Brooklyn
3. Geospatial data – latitude and longitude of each neighborhood

---

<sup>1</sup> Simone Cinotto, 2013 “*The Italian American Table, Soft Soil, Black Grapes, and Making Italian America*”

<sup>2</sup> Italian-American cuisine, Wikipedia.org

## **Sources, description of data and methods of extraction**

List of all neighborhoods of Brooklyn define scope of this project. We will use data set “2014 New York City Neighborhood Names” (Department of City Planning) from following source: [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572) which contains list of neighborhoods and boroughs of the New York city.

We will filter this data set to get only neighborhoods of Brooklyn with coordinates of each neighborhood.

Venue data, related to Italian restaurants located in Brooklyn we can find using Foursquare API, as one of the largest data provider. We will search for venue data related to Italian restaurants in neighborhoods of Brooklyn.

In this project we will use different data science skills and tools: Pandas library for data extraction, cleaning, wrangling; Foursquare API for data source; machine learning algorithm for neighborhood clustering, K-means; visualization library Folium, for neighborhood visualization on interactive map.

In Methodology section we will describe all steps taken in this project, data analysis that we performed and machine learning technique that was used.