

Coursera – IBM Data Science Professional Certification

Applied Data Science Capstone

**Opening a Authentic French Patisserie
in
New York**

By

Preetha Venkatasamy

January, 2021



Introduction

New York is the most ethnically diverse, religiously varied, commercially driven, and, in the eyes of many, the most attractive urban centre in the country. New York always meant possibility, for it was an urban centre on its way to something better, a metropolis too busy to be solicitous of those who stood in the way of progress. While people of European ancestry still make up one-third of the population, Hispanics account for nearly one-third, and African Americans about one-fourth. By constantly enhancing its key economic advantages, New York has remained prosperous even as it underwent change, its strength lying in its diversity. New York still continues to welcome many legal immigrants into the city's "golden door" to this day.

Every New York City neighborhood has one: A pastry shop where the air smells sweet, the conversation is light, and the coffee is hot. Bakeries make everything from bread to cookies, but pastry shops specialize in sugar, especially when it involves flaky doughs, frosted cakes, and glistening tarts. Open for breakfast,

pastry shops serve warm cinnamon rolls or many-layered cherry danishes, tall biscuits, or soft muffins studded with berries. Pastry shops are where diners go to order a birthday cake, meet over a slice of pie, or dive face first into a cupcake. They're almost always open all day, too, so they also attract those looking for a sweet fix just before the sun starts to set.

Business Problem

This project aims to identify the best location in New York to open a French Patisserie that would attract the culturally diverse population. This Patisserie will have Classic French offerings like tarte Bourdaloue — filled with poached pears and almond cream, Opera cake and Paris-Brest to colorful fruit tarts, chocolate eclairs and even wedding cakes. It will be open for all occasions from breakfast to dinner.

Location is one of the most important factors that will decide whether this business will be a success or a failure. With the help of Data Science methodology and tools, this project will help the business owner find the best location in New York, to start the business keeping in mind the crowd and diversity.

Data Requirement

We will need the following data to properly address the business question:

- 1) The entire New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough
- 2) The latitude and longitude coordinates for each neighborhood
- 3) Venue categories in each neighborhood mostly relevant to bakery that will be used for further analysis to decide the best location

Data Source and Extraction

New York city data containing the neighborhoods and boroughs will be obtained from the open data source: https://cocl.us/new_york_dataset. Once this data is obtained in the required format, the corresponding latitude and longitude coordinates for each neighborhood will be obtained using the python Geocoder package.

We will use the Foursquare API to explore neighborhoods in New York City. Here, we will construct a URL to send a request to the API to search for a specific type of venue that is relevant to bakery in each neighborhood, and then use this feature to group the neighborhoods into clusters. We will use the *k*-means clustering algorithm to complete this task.

Finally, we will use the Folium library to visualize the neighborhoods in New York City and their emerging clusters.