

# Applied Data Science Capstone Project

*Study of neighborhoods in New York for the launch a new Spanish food Restaurant on the “Big Apple”.*

By AWAISSSEN 2021

## 1. INTRODUCTION

New York since the end of the 19th century has been one of the centers of the world economy. It is considered a global city, due to its worldwide influences on the media, politics, education, architecture, entertainment, the arts and fashion. The great relevance of the city on many levels makes it, together with Moscow, Beijing, London, Tokyo and Paris, one of the cities with the highest statistics on the planet.

The city is made up of five boroughs: Bronx, Brooklyn, Manhattan, Queens and Staten Island (also known as "metropolitan districts" for the purposes of the city, and whose translation and administrative figure is equivalent to "municipality"), each of which coincide with a county: Bronx, Kings, New York, Queens, and Richmond. With more than 24 million New Yorkers in an urban area of 830 square kilometers (320 mi<sup>2</sup>), New York is the second most densely populated city in the United States, behind only Union City, New Jersey, located on the other side of the Hudson River.

New York has the un-official nickname for the City of Neighborhoods. There are more than 140 recognized within the city, but it is estimated at more than 240 in its total limits. Recognized too as one of the most multicultural and cosmopolitan in the world, which attracts a great number of tourists without equal within the country.

This multicultural and tourist character has its image in gastronomy, taking into account all the above, it is not surprising that it is the ideal candidate for a new restaurant serving typical Spanish food.

The objective of this project is to determine, through the study of New York neighborhoods, some possible locations for the start-up of a Spanish Restaurant. The data will be appropriately analyzed to find recommendations for project stakeholders.

## 2. DATA RESOURCES

The essential data that we are going to require for the project will be:

2.1. New York neighborhood data source

2.2. Geographical data and coordinates within New York for those neighborhoods

2.3. Data management with recommendations

### 2.1. New York neighborhood data source

The data of the neighborhoods of New York will be extracted from [https://en.wikipedia.org/wiki/Neighborhoods\\_in\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City), using **pandas data frame** with an html reading method. The reason for choosing this data source is that Wikipedia provides a detailed and understandable table that can be managed for analysis using the pandas **read\_html method**.

### 2.2. Geographical data and coordinates within New York for those neighborhoods

The geographical data for New York will be obtained through the **Python GeoPy library** and using the **Folium library** too. We obtain **latitude and longitude** data for a variety of neighborhoods in New York, to contrast with the given by Wikipedia, and replace in our data frame.

### 2.3. Data management with recommendations

The data on the final recommendations will be extracted using the **Foursquare API method**. We will essentially get the most popular neighborhoods, which would support a very good location for a new restaurant. The revenue recommendations of all neighborhoods were obtained with a limit of 150 revenue recommendations per neighborhood and a radius of 1 km around the neighborhood's geographical coordinates.

