

# IBM Data Science Professional Certificate

## Capstone Project - The Battle of Neighborhoods

Damien Azzopardi – July 2021

---

### 1. Introduction

**The Green Alternative** is a group of vegetarian restaurants, which started operating in Madrid, Spain, in 2010. We are currently running six different restaurants across different neighborhoods in Madrid, oriented towards locals. As our group is becoming successful in the Spanish capital, this year, we would like to expand our operations and open a vegetarian restaurant in Barcelona.

The question we are trying to answer is **what is the best neighborhood to open a vegetarian restaurant in Barcelona?**

After running a market research and looking into the data collected from our six current restaurants in Madrid, we found that our most successful locations are in neighborhoods which, in order of priority:

1. Are close to a **metro** or **train station**, where the flow of people is high.
2. Have a **gym** close by, as most of our customers come for lunch or dinner after training at the gym.
3. Have a **park** or **garden** close by, where our customers like to have lunch.

Knowing this, we'll leverage the Foursquare location data in order to calculate the density of metro and train stations, parks, gardens, and gyms, for each neighborhood in Barcelona, and pick the one with higher density of selected venues to open our first vegetarian restaurant in the city of Barcelona.

---

### 2. Data

The data we will be using to help us answer our question comes from the following sources.

## Neighborhoods and coordinates

### *Metabolism of Cities*

The full list of Barcelona's neighborhoods, along with their corresponding coordinates is available in [this](#) page (metabolismofcities.org). It consists of a table with two rows, **Neighborhoods** and **Coordinates**. We will scrap the table containing the list of neighborhoods and coordinates directly in this workbook.

## Districts

### *Wikipedia*

The full list of Barcelona's districts, along with their corresponding neighborhoods is available in [this](#) page (wikipedia.org). We will export a CSV containing two rows, **Districts** and **Neighborhoods**, that we will read directly in this workbook, and join it with the first dataset containing the **Neighborhoods** and **Coordinates**.

## Venues

### *Foursquare*

We will leverage the Foursquare location data in order to calculate the density of the venues we have selected for the analysis. We will join it with the first two datasets containing the **District**, **Neighborhoods** and **Coordinates**.

---

## 3. Methodology

For this analysis, we'll go for a **classification** approach using the **K Means** method. We will classify all neighborhoods in Barcelona into five different clusters, considering only the venues that match our criterion - Metro Station, Train Station, Park, Garden, Gym – and look for the clusters that are in line with the data we've collected from our six current restaurants in Madrid.

# 4. Results

Using the **K Means** method to classify all neighborhoods in Barcelona, we ended up with the following five clusters.

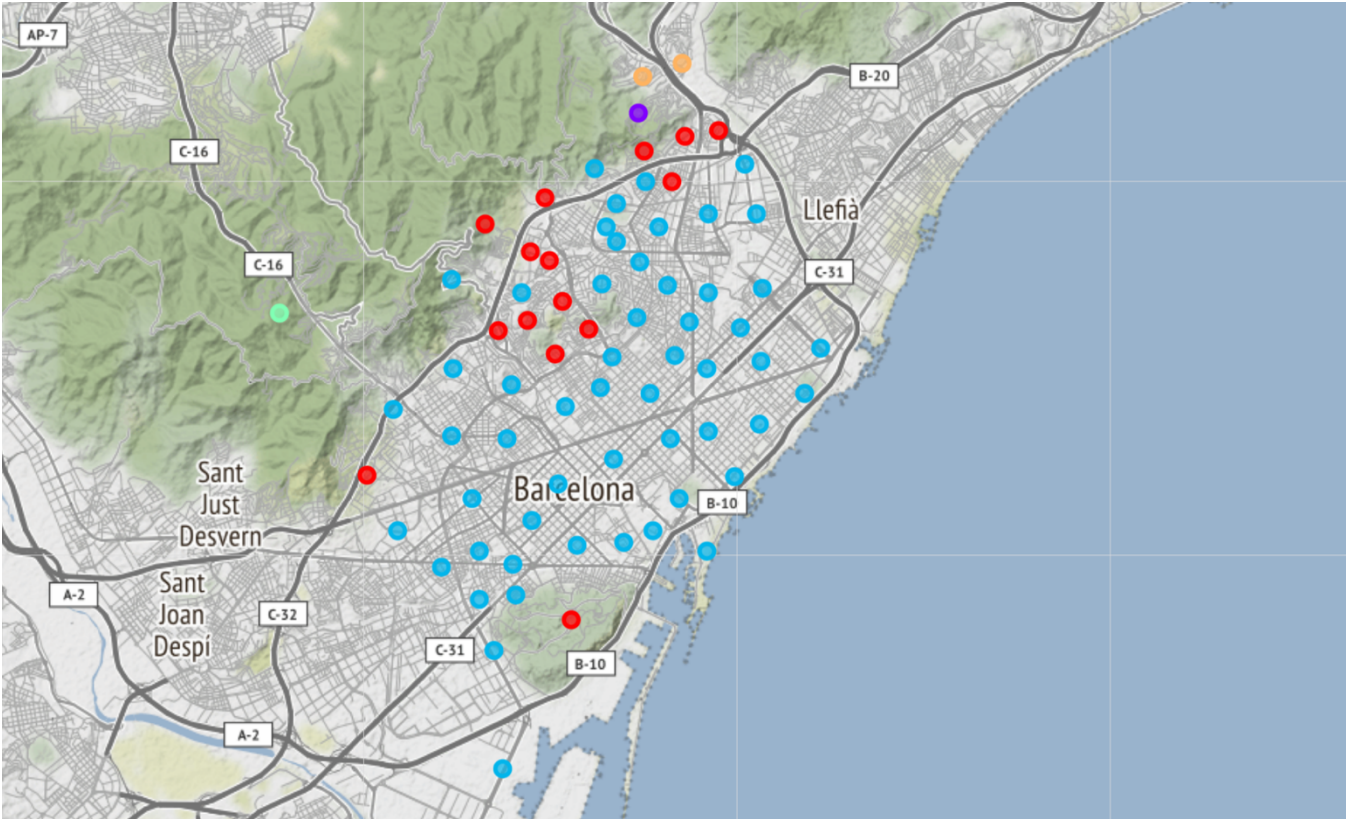


Figure 1. Map of Barcelona’s neighborhoods clustered using the K Means method

Diving into each of these clusters, we found that one exactly matches our original criterion (cluster #2, purple point on the above map), whose first most common venue is a **Metro or Train Station**, the second most common venue is a **Gym**, and the third most common venue is a **Park** or a **Garden**.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
23	Torre Baró	Metro Station	Train Station	Gym	Park	Garden

Figure 2. Cluster #2 results

---

## 5. Conclusion

Based on the data previously collected in the city of Madrid and the analysis ran on Barcelona's neighborhoods, we can safely say that the **Torre Barró** neighborhood is the best area to open a new franchise of our **Green Alternative** restaurants.

It is important to mention that this neighborhood is quite far from the city center, where most of the touristic attractions are. This should not be a problem for the **Green Alternative** as its current restaurants are oriented towards locals, and the group wants to keep this same direction. However, if the group decides to switch its focus in the future and expand its customer base to a broader audience, they will need to replicate this analysis taking into accounts new criterion, different from the ones picked for this analysis.