The battle of the neighborhoods

Finding the best location to open a restaurant in Madrid, Spain.

JANUARY 2

IBM COURSERA DATA SCIENCE CAPSTONE
Authored by: Luis Castillo

IBM

coursera

Contents

1. BUSINESS PROBLEM

Our customer, the restaurant chain “XYZ Fancy Dining” is interested in opening a new restaurant in Madrid. Madrid is one of the busiest cities in Europe, with more than three million residents and an average of almost 800.000 visitors each month.

This would be our customer’s second restaurant location, after having successfully opened a venue in Greenwich Village, a very lively neighborhood from New York city.

Considering that our customer has had very good results with their NYC location, they have requested our data science team to find a neighborhood with similar characteristics.

The problem question would be: **What neighborhood from Madrid has the most similar characteristics in terms of entertainment and dining options compared to Greenwich Village in New York City?**

2. DATA

The data to be used for this project comes from three different locations:

- Foursquare. It is a local search-and-discovery service which provides information on different types of entertainment, drinking and dining venues. Foursquare has an API that can be used to query their database and find information related to the venues, such as location, overall category, reviews and tips.
- Madrid Neighborhood Names and geographic coordinates. Available on https://datos.madrid.es/, this is used to obtain the neighborhood location information from the city.
- New York City Neighborhood Names and geographic coordinates. Data available on https://geo.nyu.edu/catalog/nyu_2451_34572

Below the details of how we will use each data source during this project.

2.1. Foursquare API data

For this project we will use the Foursquare Places API. One of the features of this API is to provide a list of venues within a specific location, based on the Lat/Lon coordinates and a radius.

In order to obtain a list of venues within a specified area, we use the “explore” endpoint from the API. By passing the proper parameters via an HTTP request to the *explore* endpoint, we get a JSON object with the information shown in the table below:

Field	Description
id	A unique string identifier for this venue.
name	The best known name for this venue.
location	An object containing none, some, or all of `address` (street address), `crossStreet`, `city`, `state`, `postalCode`, `country`, `lat`, `lng`, and `distance`. All fields are strings, except for `lat`, `lng`, and `distance`. Distance is measured in meters. Some venues have their locations intentionally hidden for privacy reasons (such as private residences). If this is the case, the parameter `isFuzzed` will be set to true, and the `lat/lng` parameters will have reduced precision.
categories	An array, possibly empty, of categories that have been applied to this venue. One of the categories will have a `primary` field indicating that it is the primary category for the venue. For the complete category tree, see categories.

Figure 1. Information contained in response to request towards "explore" endpoint

The *location* object contains the coordinates of each venue, which will be used to associate it with its respective neighborhood.
The *categories* array will be used to categorize the neighborhood. Basically, we will count how many venues from all available categories are found on each neighborhood, and then use that information to compare neighborhoods from Madrid with Greenwich Village in NYC.

2.2. Madrid Neighborhoods

The Madrid city government has made available to the public a series of datasets with information of interest. We will be using the "Divisiones administrativas: distritos, barrios y divisiones históricas" dataset, available in the following URL: https://datos.madrid.es/egob/catalogo/200078-10-distritos-barrios.zip
In order to do geographic visualizations with this data using the Folium library, it will need to be converted to JSON format. We will do this with the *geopandas* python library.

2.3. New York City Neighborhoods

The NYU has made available a dataset with the basic neighborhood information from New York City. This is available on https://geo.nyu.edu/catalog/nyu_2451_34572

3. METHODOLOGY

4. RESULTS

5. DISCUSSION

6. CONCLUSSION