# HELPING PEOPLE FIND THE BEST NEIGHBOURHOOD TO LIVE IN

# April 2021

## 1. Introduction

All over the world, people are constantly moving to different cities. They can move because of work, family, lifestyle, culture, and many other reasons.

When moving, **some people may want to live in a neighbourhood similar to where they live today**.

However, they may not know the new city they are moving to. **Therefore, we will help them find which neighbourhoods are the best match for them**, by comparing:

- Venues nearby of their current home
- Venues in each neighbourhood of the destination city

For this, we will consider a **walking range of half mile (800m)**.

Of course, there are **many other variables** that may affect this decision, such as cost of living, distance to work, criminality, etc. These **will not be included in this analysis.**

To test the models, we will run **two scenarios**:

- Scenario 1: **Jane** currently **lives in central London**, and has **received a job offer in Boston.** She doesn't know the city, and wants to know which neighbourhoods are best for her.
- Scenario 2: **John lives in New York**, and wants to live one year abroad to improve his Spanish. **He decided to go to Madrid. However**, to reduce the culture shock, he wants to live in a neighbourhood similar to the one he currently lives (or as similar as possible)

# 2. Data

## 2.1. Sources

- **Boston neighborhoods**: The initial list used was from Boston's government website https://www.boston.gov/neighborhoods
    - However, later we decided to **adopt a 'grid' approach for better precision**, due to low number of neighbourhoods. This didn't require any data sources
- **Madrid neighborhoods**: The list used was from Wikipedia https://en.wikipedia.org/wiki/List_of_neighborhoods_of_Madrid
- Current **addresses** from our 'clients'
    - Jane lives in Rushworth St, London
        - Latitude: 51.501463
        - Longitude: -0.1020907
    - John lives in Scholes St, Brooklyn, NY
        - Latitude: 40.708179
        - Longitude: -73.949628

## 2.2. APIs and libraries used

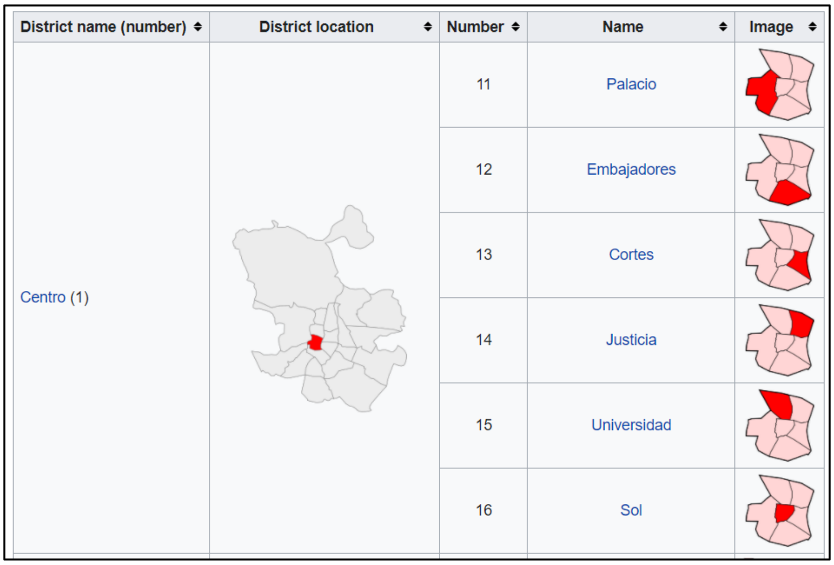
- **Geopy**: used to get latitude and longitude from addresses (and addresses from latitude and longitude)
- **Foursquare**: used to fetch nearby venues based on latitude and longitude
- **Folium**: used to generate maps
- **Pandas**: used to work with dataframes
- **Numpy**: used for simple numeric processing
- **Requests**: used to get html data from websites
- **BeautifulSoup**: used to process html data

## 2.3. Pre-processing

The data from the websites was processed into dataframes using **BeautifulSoup and Pandas.**

The **Boston data was well structured**, but it was later discarded as we chose another better approach.

The Madrid data was not well structured, and **required pre-processing** to achieve the desired results. The specific code can be found in the notebook.



| District name (number) | District location | Number | Name | Image |
|---|---|---|---|---|
| Centro (1) | | 11 | Palacio | |
| | | 12 | Embajadores | |
| | | 13 | Cortes | |
| | | 14 | Justicia | |
| | | 15 | Universidad | |
| | | 16 | Sol | |

Example of Madrid neighbourhood data from Wikipedia

# 3. Methodology

The methodology consisted of six steps. Results from these is in the 'Results' section

1. Use **HTML processing** (requests and BeautifulSoup) to get the **list of neighborhoods of each destination city into dataframes**
2. Use APIs to get the **latitude and longitude of each neighborhood**
3. For each location, get a **list of nearby venues** and their types (e.g. restaurants, bars, nightclubs, etc) using the **Foursquare API**, and **consolidate by neighborhood**
4. **Repeat** the process for each of our **clients' current addresses**
5. Use some measure of **similarity to identify the best neighborhoods** for our clients(e.g. Euclidean distance)
6. **Analyze the results** and plot on a map using the Folium library

# 4. Results

## 4.1. HTML Processing

The initial neighbourhood list for Boston consisted of 24 neighbourhoods, listed and plotted below.

| | Neighborhood |
|---|---|
| 0 | Allston |
| 1 | Back Bay |
| 2 | Bay Village |
| 3 | Beacon Hill |
| 4 | Brighton |
| 5 | Charlestown |
| 6 | Chinatown |
| 7 | Dorchester |
| 8 | Downtown |
| 9 | East Boston |
| 10 | Fenway-Kenmore |
| 11 | Hyde Park |
| 12 | Jamaica Plain |
| 13 | Mattapan |
| 14 | Mid-Dorchester |
| 15 | Mission Hill |
| 16 | North End |
| 17 | Roslindale |
| 18 | Roxbury |
| 19 | South Boston |
| 20 | South End |
| 21 | West End |
| 22 | West Roxbury |
| 23 | Wharf District |

Boston Neighbourhoods

These were later changed (details in section 4.2)

For Madrid, our list contained 131 neighbourhoods across 21 districts. The district data was used only to facilitate the lagitude and longitude collection. All other analysis used only the neighbourhood.

| | District | Neighborhood |
|---|---|---|
| 0 | Centro | Palacio |
| 1 | Centro | Embajadores |
| 2 | Centro | Cortes |
| 3 | Centro | Justicia |
| 4 | Centro | Universidad |
| ... | ... | ... |
| 126 | Barajas | Alameda de Osuna |
| 127 | Barajas | Aeropuerto |
| 128 | Barajas | Calle Canal de Suez |
| 129 | Barajas | Timón |
| 130 | Barajas | Corralejos |

Sample Madrid neighbourhoods

## 4.2. Latitudes and Longitudes

Latitudes and longitudes were extracted using geopy's Nominatim feature. Here's a sample of how you can use it:

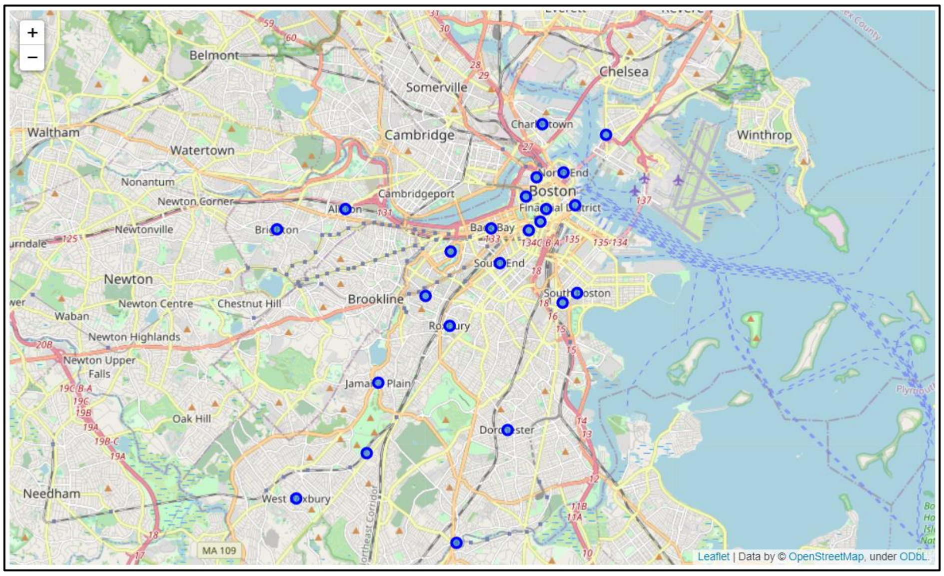
```
geolocator = Nominatim(user_agent="coursera-capstone")
location = geolocator.geocode("Eiffel Tower, Paris, France")
print("Location of the Eiffel Tower: Latitude = {}, Longitude = {}".format(location.latitude, location.longitude))

Location of the Eiffel Tower: Latitude = 48.858260200000004, Longitude = 2.2944990543196795
```

geopy's Nominatim example

The results from Boston were plotted using Folium.

| | Neighborhood | Latitude | Longitude |
|---|---|---|---|
| 0 | Allston | 42.355434 | -71.132127 |
| 1 | Back Bay | 42.350549 | -71.080311 |
| 2 | Bay Village | 42.350011 | -71.066948 |
| 3 | Beacon Hill | 42.358708 | -71.067829 |
| 4 | Brighton | 42.350097 | -71.156442 |
| 5 | Charlestown | 42.377875 | -71.061996 |
| 6 | Chinatown | 42.352217 | -71.062607 |
| 7 | Dorchester | 42.297320 | -71.074495 |
| 8 | Downtown | 42.355431 | -71.060500 |
| 9 | East Boston | 42.375097 | -71.039217 |
| 10 | Fenway-Kenmore | 42.344224 | -71.094444 |
| 11 | Hyde Park | 42.255654 | -71.124496 |
| 12 | Jamaica Plain | 42.309820 | -71.120330 |
| 13 | Mattapan | 42.267566 | -71.092427 |
| 14 | Mid-Dorchester | 42.330786 | -71.054750 |
| 15 | Mission Hill | 42.332560 | -71.103608 |
| 16 | North End | 42.365097 | -71.054495 |
| 17 | Roslindale | 42.291209 | -71.124497 |
| 18 | Roxbury | 42.324843 | -71.095016 |
| 19 | South Boston | 42.333431 | -71.049495 |
| 20 | South End | 42.341310 | -71.077230 |
| 21 | West End | 42.363919 | -71.063899 |
| 22 | West Roxbury | 42.279265 | -71.149497 |
| 23 | Wharf District | 42.356447 | -71.050324 |

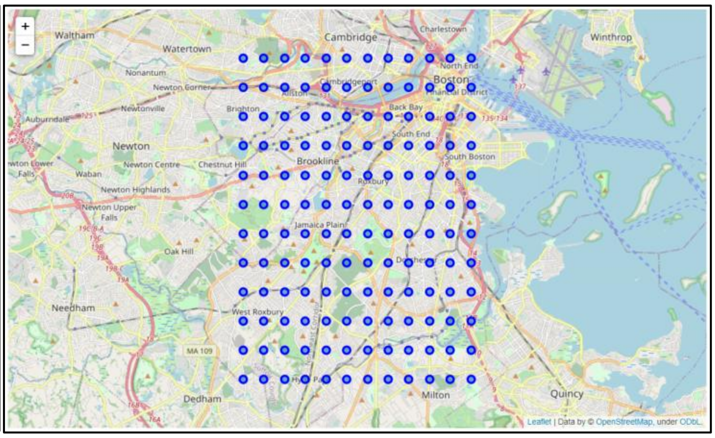



Boston neighbourhoods latitude and longitude

It is clear that these are very far apart and would not lead to great results. **Therefore, we adopted a grid approach**, which can be seen below. These contain more datapoints and will lead to better results.

We chose a 12x12 grid, leading to 144 points, similar to Madrid's 131.

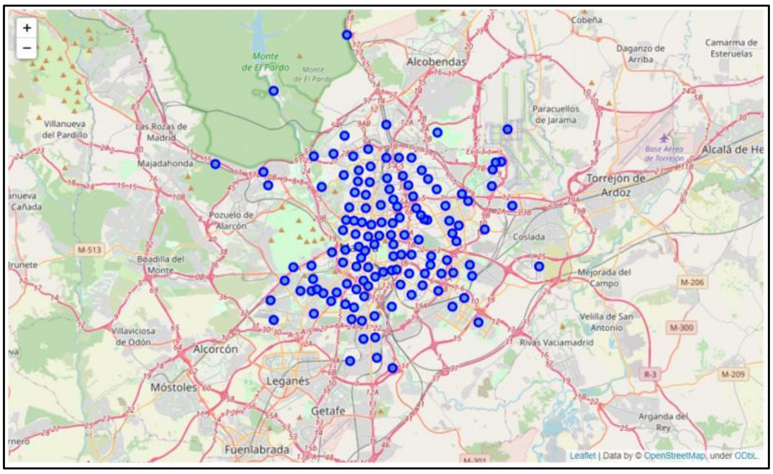
| | Latitude | Longitude | Neighborhood |
|---|---|---|---|
| 0 | 42.255654 | -71.156442 | 121, Rockland Street, East Dedham, Dedham, Nor... |
| 1 | 42.255654 | -71.146674 | 6, Hyde Park Street, Dedham, Norfolk County, M... |
| 2 | 42.255654 | -71.136905 | Stony Brook Reservation, Dedham Boulevard, Eas... |
| 3 | 42.255654 | -71.127136 | 1339, River Street, Hyde Park, Boston, Suffolk... |
| 4 | 42.255654 | -71.117367 | Fairmount, Walter Street, Fairmount, Hyde Park... |
| ... | ... | ... | ... |
| 139 | 42.367690 | -71.088061 | 152, Fulkerson Street, East Cambridge, Cambrid... |
| 140 | 42.367690 | -71.078292 | 16, Hurley Street, East Cambridge, Cambridge, ... |
| 141 | 42.367690 | -71.068524 | Craigie Drawbridge, Charles River Dam Road, We... |
| 142 | 42.367690 | -71.058755 | North Washington Street Bridge Replacement (20... |
| 143 | 42.367690 | -71.048986 | Boston HarborWalk, Dock Square, North End, Bos... |




Boston datapoints latitude and longitude

For Madrid, the neighbourhood list **already provided us the granularity that we needed**, so we did not need to adopt the grid approach.

| | District | Neighborhood | Latitude | Longitude |
|---|---|---|---|---|
| **0** | Centro | Palacio | 40.415129 | -3.715618 |
| **1** | Centro | Embajadores | 40.409681 | -3.701644 |
| **2** | Centro | Cortes | 40.414348 | -3.698525 |
| **3** | Centro | Justicia | 40.423957 | -3.695747 |
| **4** | Centro | Universidad | 40.425310 | -3.706630 |
| **...** | ... | ... | ... | ... |
| **126** | Barajas | Alameda de Osuna | 40.457581 | -3.587975 |
| **127** | Barajas | Aeropuerto | 40.494838 | -3.574081 |
| **128** | Barajas | Calle Canal de Suez | 40.473524 | -3.579216 |
| **129** | Barajas | Timón | 40.473171 | -3.584152 |
| **130** | Barajas | Corralejos | 40.468164 | -3.587073 |




Madrid neighbourhoods latitude and longitude

## 4.3. Nearby venues (destination cities)

The neighbourhood venues were collected and then compiled by neighbourhood, since we are not interested in specific venues, but the overall trend in each neighbourhood.

The free version of the API limits calls to 100 venues. Some neighbourhoods have the maximum. But others, presumably more isolated, do not reach that. Here is the count for some neighbourhoods in Madrid.

| Neighborhood | Neighborhood Latitude |
|---|---|
| **Abrantes** | 16 |
| **Acacias** | 100 |
| **Adelfas** | 94 |
| **Aeropuerto** | 15 |
| **Alameda de Osuna** | 36 |
| **...** | ... |
| **Ventas** | 19 |
| **Villaverde Alto** | 7 |
| **Vinateros** | 23 |
| **Vista Alegre** | 30 |
| **Zofío** | 16 |

Number of venues in Madrid neighbourhoods

We then used one-hot encoding for each neighbourhood (values represented here are zero, but they are not positive in other columns).

| | NeighborhoodName | Accessories Store | Airport | Airport Lounge | Airport Service | American Restaurant | Aquarium | Arcade | Argentinian Restaurant | Art Gallery | ... | Video Game Store | Vietnamese Restaurant | Warehouse Store | Watch Shop | Whisky Bar | Wine Bar | Wine Shop | Women's Store |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Abrantes | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **1** | Acacias | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **2** | Adelfas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **3** | Aeropuerto | 0 | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **4** | Alameda de Osuna | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **126** | Ventas | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **127** | Villaverde Alto | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **128** | Vinateros | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **129** | Vista Alegre | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **130** | Zofío | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

One-hot encoding for Madrid

Finally, we aggregated the top 10 venues per neighbourhood.

| | NeighborhoodName | Top 1 | Top 2 | Top 3 | Top 4 | Top 5 | Top 6 | Top 7 | Top 8 | Top 9 | Top 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Abrantes | Metro Station | Plaza | Ice Cream Shop | Athletics & Sports | Burger Joint | Nightclub | Tapas Restaurant | Fast Food Restaurant | Park | Gym / Fitness Center |
| 1 | Acacias | Bar | Tapas Restaurant | Spanish Restaurant | Coffee Shop | Pizza Place | Art Gallery | Plaza | Indie Theater | Vegetarian / Vegan Restaurant | Market |
| 2 | Adelfas | Spanish Restaurant | Grocery Store | Bar | Bakery | Fast Food Restaurant | Gym | Burger Joint | Pizza Place | Supermarket | Hotel |
| 3 | Aeropuerto | Airport Lounge | Spanish Restaurant | Coffee Shop | Duty-free Shop | Sporting Goods Shop | Fast Food Restaurant | Breakfast Spot | Airport Service | Diner | French Restaurant |
| 4 | Alameda de Osuna | Restaurant | Hotel | Park | Spanish Restaurant | Hotel Bar | Café | Gym | Tapas Restaurant | Bistro | Coffee Shop |

Top 10 venues in Madrid neighbourhoods

## 4.4. Nearby venues (current client homes)

Using the same process as above, we consolidated the results around Jane's and John's current houses.

| | NeighborhoodName | American Restaurant | Argentinian Restaurant | Art Gallery | Art Museum | Arts & Crafts Store | Asian Restaurant | Bagel Shop | Bakery | Bar | ... | Taiwanese Restaurant | Thai Restaurant | Theater | Thrift / Vintage Store |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Jane's neighborhood | 1 | 2 | 1 | 3 | 0 | 1 | 0 | 1 | 2 | ... | 0 | 0 | 3 | 0 |
| 1 | John's neighborhood | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 4 | 9 | ... | 1 | 2 | 0 | 1 |

Jane and John's nearby venues by category

And we also got the top 10 for each of them.

| | NeighborhoodName | Top 1 | Top 2 | Top 3 | Top 4 | Top 5 | Top 6 | Top 7 | Top 8 | Top 9 | Top 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Jane's neighborhood | Coffee Shop | Hotel | Pub | Gym / Fitness Center | Café | Theater | Italian Restaurant | Art Museum | Portuguese Restaurant | Argentinian Restaurant |
| 1 | John's neighborhood | Bar | Pizza Place | Coffee Shop | Italian Restaurant | Bakery | Japanese Restaurant | Latin American Restaurant | Wine Shop | Restaurant | Food Truck |

Top 10 venue categories for Jane and John

**We can see that Jane lives next to coffee shops, hotels, pubs and gyms. John lives next to bars, pizza places, coffee shops and Italian restaurants.**

## 4.5. Compare neighbourhoods

Neighbourhoods were compared using Euclidean distance across all venue categories.

First we compared Jane's current address to all of Boston's datapoints. **The maximum 'distance' between Jane and Boston's neighbourhoods is 25.2and the minimum is 14.4.** Remember that distance, in this case, is the opposite of similarity between neighbourhoods.

| | Neighborhood | Distance |
|---|---|---|
| 0 | 10, Wooddale Avenue, Mattapan, Boston, Suffolk... | 19.000000 |
| 1 | 1000, Harvard Street, Boston, Suffolk County, ... | 19.339080 |
| 2 | 104, Reedsdale Road, Milton Center, Milton, No... | 19.235384 |
| 3 | 1084, Boylston Street, Back Bay, Boston, Suffo... | 15.842980 |
| 4 | 11, Norway Road, Milton Upper Mills, Milton, N... | 18.920888 |
| ... | ... | ... |
| 139 | The Jewish Advocate, School Street, Downtown C... | 14.456832 |
| 140 | United States Postal Service Lot A, A Street, ... | 15.297059 |
| 141 | Untitled Landscape, Boston HarborWalk, Waterfr... | 16.673332 |
| 142 | Walter C. Wood Sailing Pavilion, 134, Memorial... | 16.000000 |
| 143 | Williams Street, Jamaica Plain, Boston, Suffol... | 16.852300 |

Euclidean distance between venues on Jane's neighbourhood and Boston

We repeated the same for John and Madrid. **The maximum 'distance' between John and Madrid's neighbourhoods is 26.2 and the minimum is 12.5.** Remember that distance, in this case, is the opposite of similarity between neighbourhoods

| | Neighborhood | Distance |
|---|---|---|
| **0** | Abrantes | 15.905974 |
| **1** | Acacias | 12.529964 |
| **2** | Adelfas | 17.262677 |
| **3** | Aeropuerto | 16.462078 |
| **4** | Alameda de Osuna | 16.062378 |
| **...** | ... | ... |
| **126** | Ventas | 15.459625 |
| **127** | Villaverde Alto | 16.401219 |
| **128** | Vinateros | 14.899664 |
| **129** | Vista Alegre | 14.933185 |
| **130** | Zofío | 17.146428 |

Euclidean distance between venues on John's neighbourhood and Madrid

Finally, we used the distance to **calculate a 'score'** measure between 0 and 1. The smaller the distance, the better the score.

| | Neighborhood | Distance | Score |
|---|---|---|---|
| **0** | Abrantes | 15.905974 | 0.753229 |
| **1** | Acacias | 12.529964 | 1.000000 |
| **2** | Adelfas | 17.262677 | 0.654060 |
| **3** | Aeropuerto | 16.462078 | 0.712580 |
| **4** | Alameda de Osuna | 16.062378 | 0.741796 |

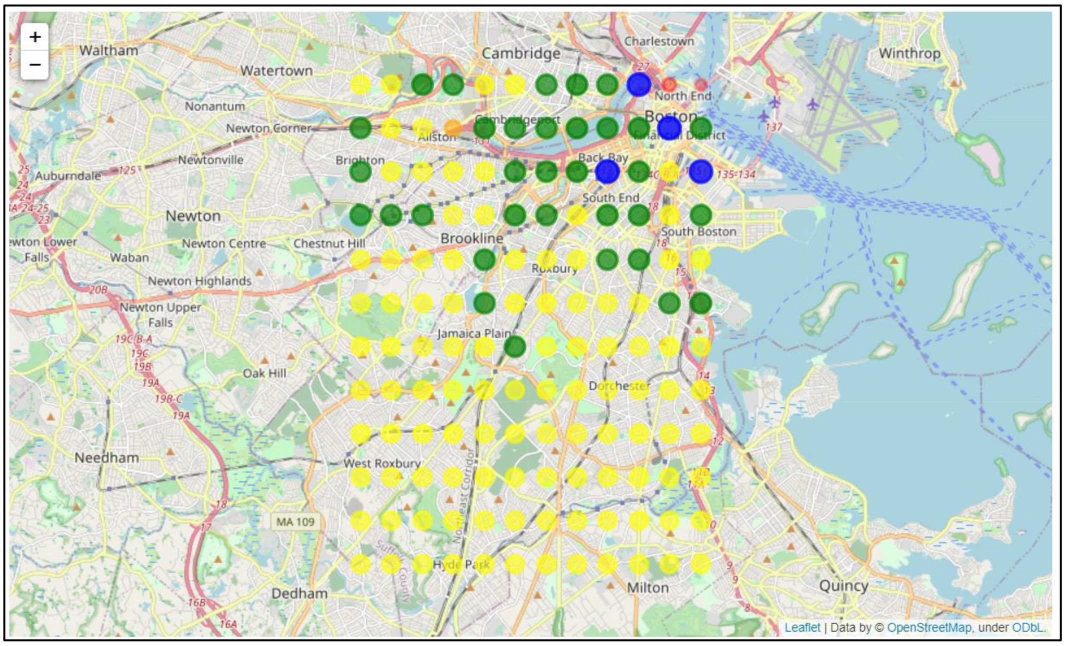
Madrid neighbourhood scored for John

## 4.6. Analyse and plot results

A **colour code** was used when plotting the results, to facilitate visualization, using the following criteria:

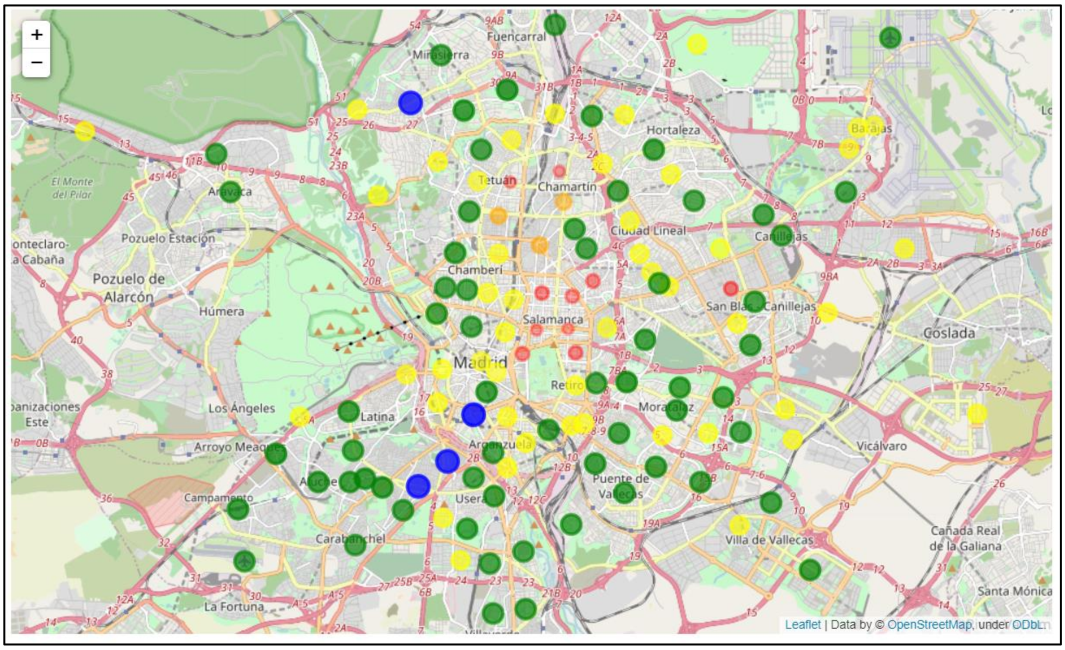
- Score above 0.9: blue
- Score between 0.7 and 0.9: green
- Score between 0.5 and 0.7: yellow
- Score between 0.3 and 0.5: orange
- Score below 0.3: red

First, we mapped **Boston's neighbourhoods based on their compatibility with Jane** (the results will be discussed in the next session).




Map of Boston’s neighbourhoods based on compatibility with Jane

We did the same for **Madrid and John**.




Map of Madrid’s neighbourhoods based on compatibility with John

# 5. Discussion and recommendations

We were **successfully** able to find the **best neighbourhoods** for Jane in Boston, and for John in Madrid.

For Jane, we recommend her to move **somewhere close to city center**, especially **Downtown Crossing, Back Bay or Seaport District**. However, there are many other places in Boston where she would find herself at home.

For John, we **do not recommend the city center**, especially not around Salamanca. He should move to neighbourhoods **slightly towards the outskirts, such as Acacias, Opañel and Comillas.**

# 6. Conclusion

This was a very interesting project. **We were able to find the best fit for two different people in two different cities.** Just as important, we also saw **which neighbourhoods are not a good fit** for them.

As we mentioned in the beginning, venue similarity is only one of multiple factors that should impact people's choice of neighbourhood to live in. Other factors for further analysis and studies are recommended in the 'Next steps' below.

## 6.1. Next possible steps and further analysis

- **Zoom in** on selected regions to find the best streets or blocks within neighbourhoods
- Experiment with **different clients** moving to Boston and Madrid
- Experiment with **different cities**
- Incorporate **distance to workplace** into analysis
- Incorporate **criminality levels** into analysis
- Incorporate **cost of living** into analysis
- Incorporate **personal inputs from clients** into analysis