



# Madrid Real Estate Experts

**Finding you the right place...**

# Business Problem:

A real state company has an important client that lives in a district located in the North of Madrid, and she asks to be advised about the most similar neighbourhood in the South of Spain's capital, as she is planning to move closer to her job place but loves very much her neighbourhood

The client tells us she is living in "Tetuan" District and she would like to move to one of the following neighbourhoods down south: Moratalaz, Retiro, Arganzuela

So, considering the big amount of data and venues available in Foursquare, we will check statistically which one of the three neighbourhoods in the south of Madrid has similar amenities to "Tetuan" in the north, so we can propose a suitable area

# Describing Data Available

Given the option to access the venues placed in each of these three neighborhoods using Foursquare API, we will scrap the data and analyse which kind of venues we have available in our currently and preferred neighborhood (Tetuán) and the candidate neighborhoods (Moratalaz, Retiro and Arganzuela).

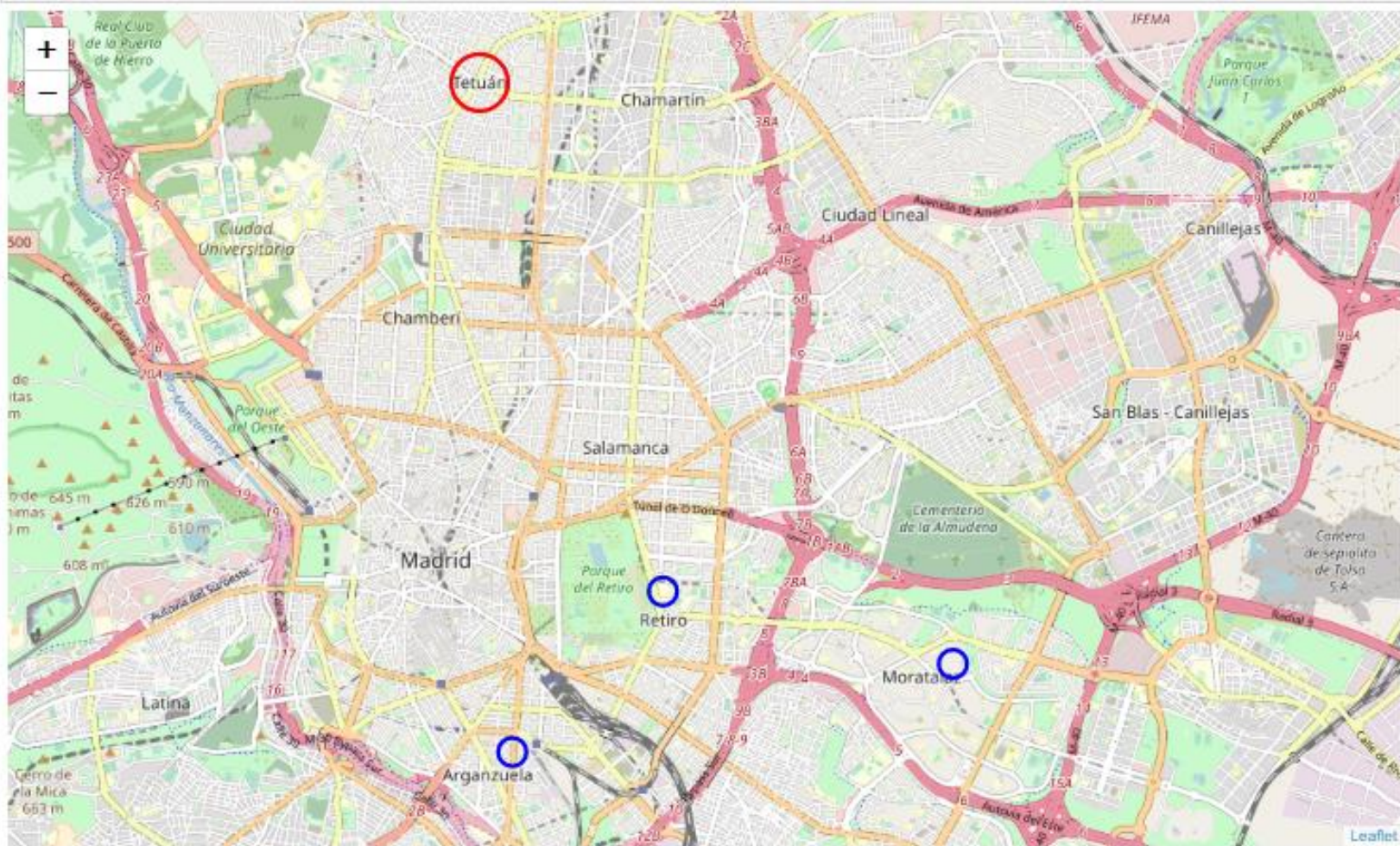
Once we have the data, we will define a Content Based Recommendation model. ¿What are the steps to make that posible?

- Defining the 4 Neighborhoods and their coordinates in a pandas dataframe
- Getting the venues data with FourSquare API and storing them for each different Neighborhood
- Processing the data:
  1. Getting dummy variables from column "Category of venue"
  2. Calculating frequency of occurrence
  3. Assuming our client gives maximum rating to her current Neighborhood
  4. Based on the amenities her current Neighborhood has, calculating which of the three given areas is a similar choice



# Locating the Neighborhoods

Out[13]:



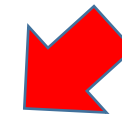
Our client lives in the **red circle** and she is used to its amenities, so let's check which one of the **three blue destination neighbourhoods** will look as familiar as possible.

# Which data to use

```
In [18]: mad_venues.head()
```

```
Out[18]:
```

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Tetuán	40.4605	-3.6982	Asador Donostiarra	40.461761	-3.695967	Spanish Restaurant
1	Tetuán	40.4605	-3.6982	El Jamón y el Churrasco	40.460516	-3.696212	Spanish Restaurant
2	Tetuán	40.4605	-3.6982	Ducati Madrid	40.462315	-3.696406	Motorcycle Shop
3	Tetuán	40.4605	-3.6982	Sabor Gaucho	40.460190	-3.694397	Brazilian Restaurant
4	Tetuán	40.4605	-3.6982	La Papita	40.459668	-3.698950	Bar



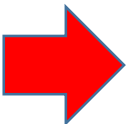
Venue Category is the column in which we can rely to tell us how a neighborhood is. That is to say that you like your neighborhood for its amenities or options that you have nearby on a daily basis.

These values can be treated as ítems to build a **content based recommendation engine**, so we can compare Neighborhoods based on their content or amenities.

# Results

Out[82]:

	Similarity	Neighborhood
2	0.047000	Retiro
0	0.028100	Arganzuela
1	0.023429	Moratalaz



As we can see in the previous table, the most similar neighborhood to "Tetuán", the place where our client currently lives, is the Neighborhood of Retiro.

To make sure about that, lets review top 10 most common places in each neighborhood:



Out[83]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arganzuela	Restaurant	Tapas Restaurant	Spanish Restaurant	Art Gallery	Indie Theater	Gym	Bakery	Grocery Store	Garden	Coffee Shop
1	Moratalaz	Bar	Spanish Restaurant	Plaza	Park	Pizza Place	Playground	Café	Soccer Field	Bakery	Diner
2	Retiro	Spanish Restaurant	Tapas Restaurant	Hotel	Restaurant	Italian Restaurant	Garden	Mediterranean Restaurant	Coffee Shop	Brewery	Seafood Restaurant
3	Tetuán	Spanish Restaurant	Restaurant	Hotel	Italian Restaurant	Tapas Restaurant	Japanese Restaurant	Pub	Burger Joint	Chinese Restaurant	Ice Cream Shop

As we can see, Retiro and Tetuan has lots of places to wine and dine, which might be the main preference of our client, while Arganzuela and Moratalaz have a bigger amount of amenities house-related such as Gyms, Bakeries, Grocery Stores, Playground and so on.

The output of the model seems to be coherent with reality, given that the non-chosen target Neighborhoods are residential areas, while Retiro and Tetuan are near the comercial and business area of Madrid, placed around the center.