



The Battle of Neighborhoods

“The scent of papyrus” (the Presentation)

Angel Ramón Ampudia Barbudo, July the 23rd of 2020

Background.

- These are not good years for business and trading with traditional paper books is not in a much healthier situation.
- The crisis not only affects stores, it is also causing economic distress to the editors, the publishers, the printers and all the components of a industry that surrounds and, in some aspect, still relies on the historical conception of books.

Background.

The time to reinvent ourselves has come, is time to raise the bet.

- Most of the bookstores are out of date, this is widely known, their economic strategy should be reformulated from the bottom-up.
- It's time for creativity.

Background.

The readers and their “nostalgia for paper”.

- "Once a reader is born it lives as a reader all its entire life, then the books turn into a basic necessity, and into the greatest common divisor of its life".
- Far from this idealistic concept, the science, and being more accurate, the statistic, puts our feet on a more solid terrain: there are ages and ages for reading.

Background.

Culture attracts culture.

- Bookstores develop in a faster manner in areas where the presence of culture is notorious.
- Monumental and cultural areas makes bookstores proliferate.
- The proximity to schools, colleges, universities and other types of centers of study pave the way to bookstores, students need books and that is good for the business.

Background.

Free is better than cheap.

- Like in many other countries, in Spain books are accessible for free and books are available to anybody with an identification card or a passport, after the registration process every user has the possibility of bringing the books home using the loan system.
- Public libraries are almost everywhere, most of the neighborhoods in Madrid have one.

Data.

List of Boroughs and Neighborhoods of Madrid

- Extracted from the Wikipedia site.
- This set of data will let us position each neighborhood of Madrid in the map.

Data.

Population by Age and Neighborhood in Madrid

- Extracted from the official site of the city council of Madrid.
- The population age will let us discriminate initially which neighborhood should be considered in the study and which not.

Data.

List of Venues by Neighborhood in Madrid

- Extracted from Foursquare.
- Some of the categories obtained from its API are:
 - School
 - College & University
 - Art Gallery
 - Art Museum
 - History Museum
 - Science Museum
 - Historic Site
 - ...

Data.

List of Public Libraries by Neighborhood in Madrid

- Extracted from the official page of the city council of Madrid.
- A public library near our local could have repercussions in the viability of our adventure. If possible, neighborhoods with public libraries should be excluded from the analysis.

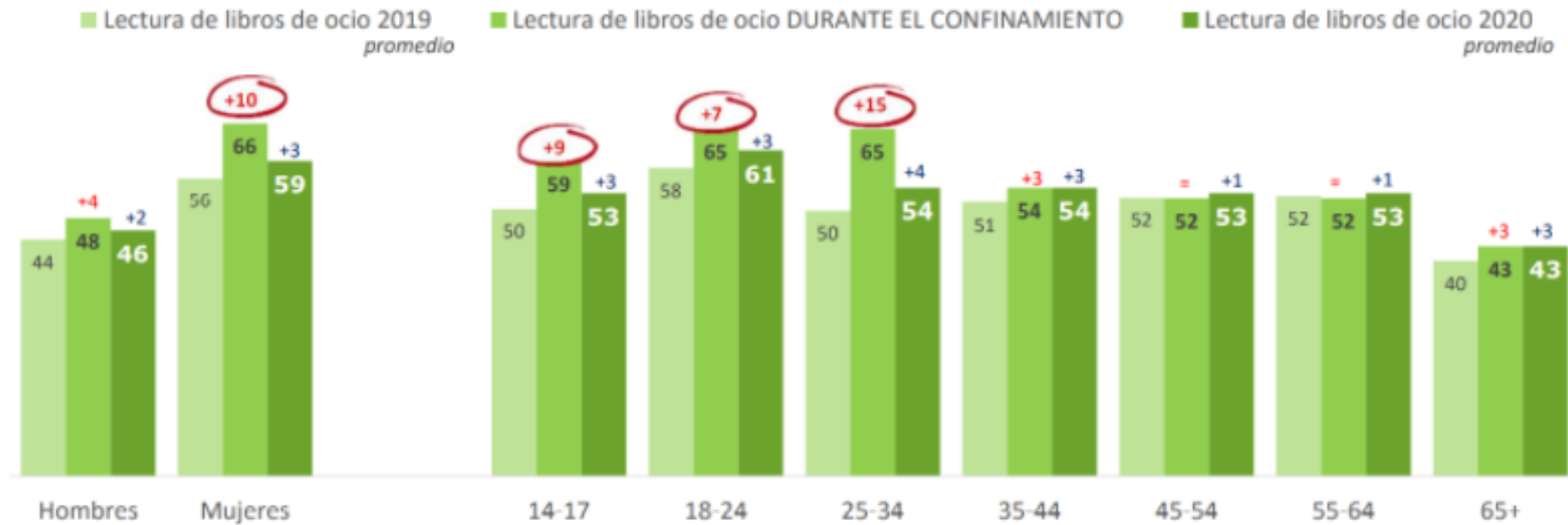
Metodology.

Reducing the data.

- There are 131 neighborhoods in Madrid, for the sake of clarity, an initial discarding process will be applied.
- The Federation of Spanish Editors recently published a report describing the reading habits of the Spanish population, positioning the peak in the ages between 18 and 44.
- Placing the threshold in the 40% of population with an age between 18 and 44 years, the list of neighborhoods is reduced to 14 elements, this will speed up the analysis of the remaining neighborhoods.

Metodology.

Reducing the data.



Metodology.

Discovering what's around with the eyes of Foursquare.

- Foursquare is a geolocation service oriented to social networking. Through the use of its API, Foursquare returns the venues situated geographically close to a given point in a radius of influence.
- In this case the categories needed are those that somehow promote the proliferation of bookstores, in other words, those that are connected to culture, teaching and architectural monuments.

Metodology.

Brotherhood of neighborhoods using machine learning.

- Through the use of machine learning computers can learn to do some labor improving their methods by their own experience.
- Some machine learning algorithms allows the association of elements considering their attributes, also described as "features", each of these groups called "clusters" includes elements that, for a given reason, are similar or compatible between them, the mechanism of grouping these elements are called "clustering".
- K-Means is an unsupervised algorithm, that indicates that the model utilized does not need to be supervised by any user, just the opposite, the algorithm allows the model to work by itself to discover information unnoticed previously.

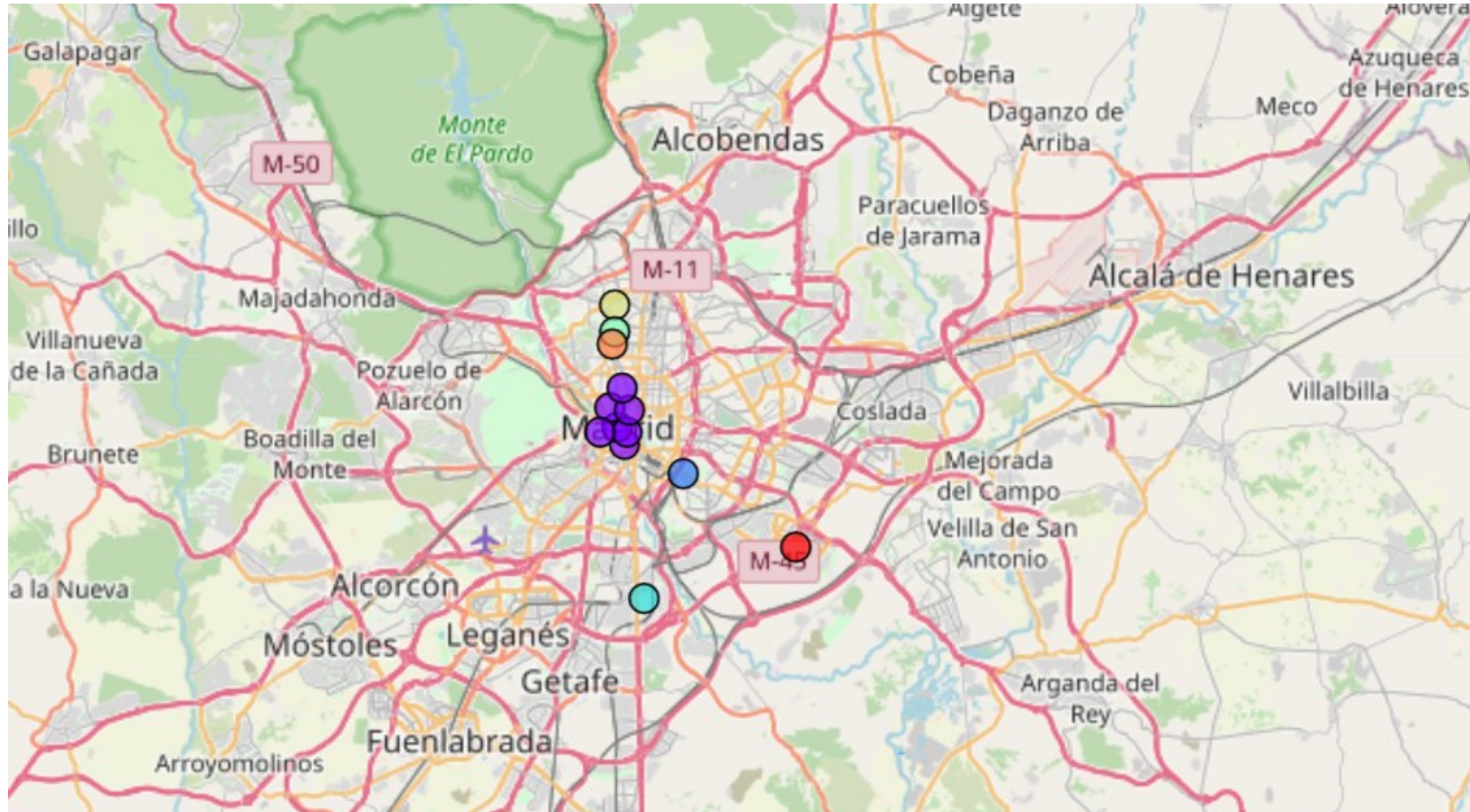
Metodology.

Brotherhood of neighborhoods using machine learning.

- The next figure displays the configuration of the seven clusters obtained by the machine learning algorithm, most of them with only one item associated and only one cluster conformed with seven items:

Metodology.

Brotherhood of neighborhoods using machine learning.



Metodology.

Avoiding unfair (fair) competence.

- Madrid has 32 public libraries integrated in its municipal cultural network.
- The access to culture should be universal and free, that is a fact nobody question in Spain, but public libraries can affect the selling rates of a bookstore in a negative way, therefore is also a fact that public libraries should be taken in consideration when looking for a place to start a business.

Metodology.

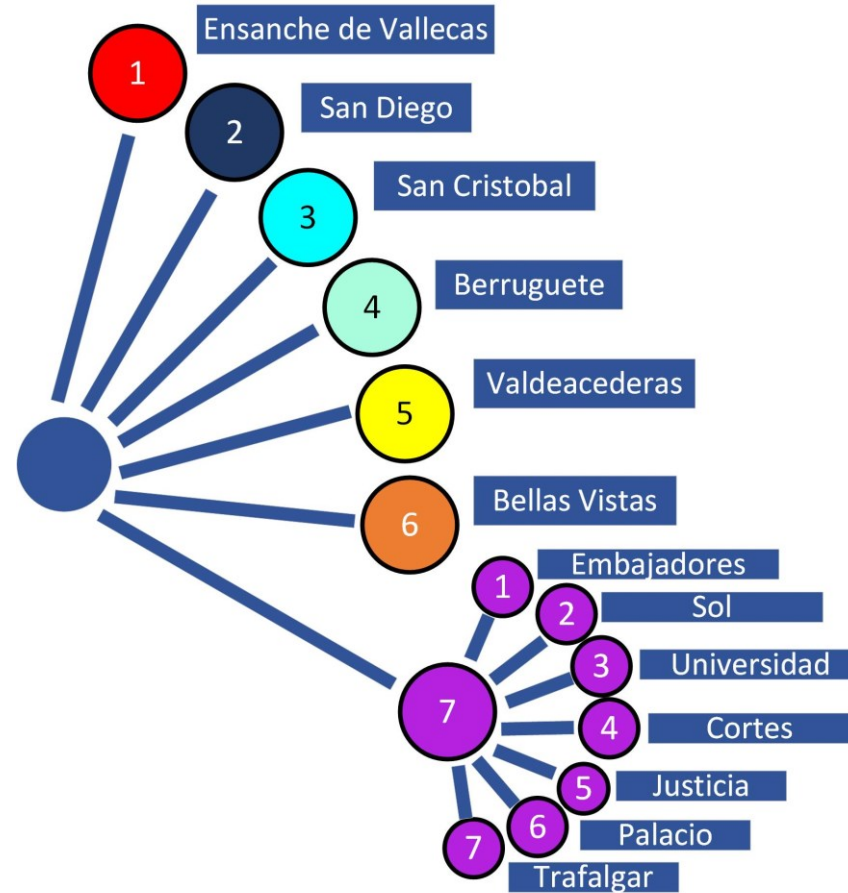
Best price is the best prize.

- Estate agencies proliferate nowadays, so is an easy task compelling a fairly good amount of economical proposals to evaluate, for our study and for simplicity purposes, all the information will be obtained from idealista.com.
- Today's statistics reveals that in Madrid and in this year of 2020, commercial premises are cheaper in the city center than in the periphery.

Results.

- The clusters determined by the machine learning algorithm are 7, the total amount of neighborhoods that has been distributed in those clusters is 13:

Results.



Results.

Cluster 1 - Ensanche de Vallecas (Villa de Vallecas).

- **The 1st Most Common Venue is:**
 - Driving School 1
- **The 2nd Most Common Venue is:**
 - Middle School 1

Results.

Cluster 2 - San Diego (Puente de Vallecas).

- **The 1st Most Common Venue is:**
 - Performing Arts Venue 1
- **The 2nd Most Common Venue is:**
 - Middle School 1

Results.

Cluster 3 - San Cristobal (Villaverde).

- **The 1st Most Common Venue is:**
 - School 1
- **The 2nd Most Common Venue is:**
 - College Library 1

Results.

Cluster 4 - Berruguete (Tetuán).

- **The 1st Most Common Venue is:**
 - Dance Studio 1
- **The 2nd Most Common Venue is:**
 - Performing Arts Venue 1

Results.

Cluster 5 - Valdeacederas (Tetuán).

- **The 1st Most Common Venue is:**
 - Student Center 1
- **The 2nd Most Common Venue:**
 - Art Gallery 1

Results.

Cluster 6 - Bellas Vistas (Tetuán).

- **The 1st Most Common Venue is:**
 - Performing Arts Venue 1
- **The 2nd Most Common Venue is:**
 - Coworking Space 1

Results.

Cluster 7 - Embajadores, Sol, Universidad, Cortes, Justicia, Palacio and Trafalgar.

- **The 1st Most Common Venues are:**

- Theater 4
- Historic Site 1
- Art Gallery 1
- Student Center 1

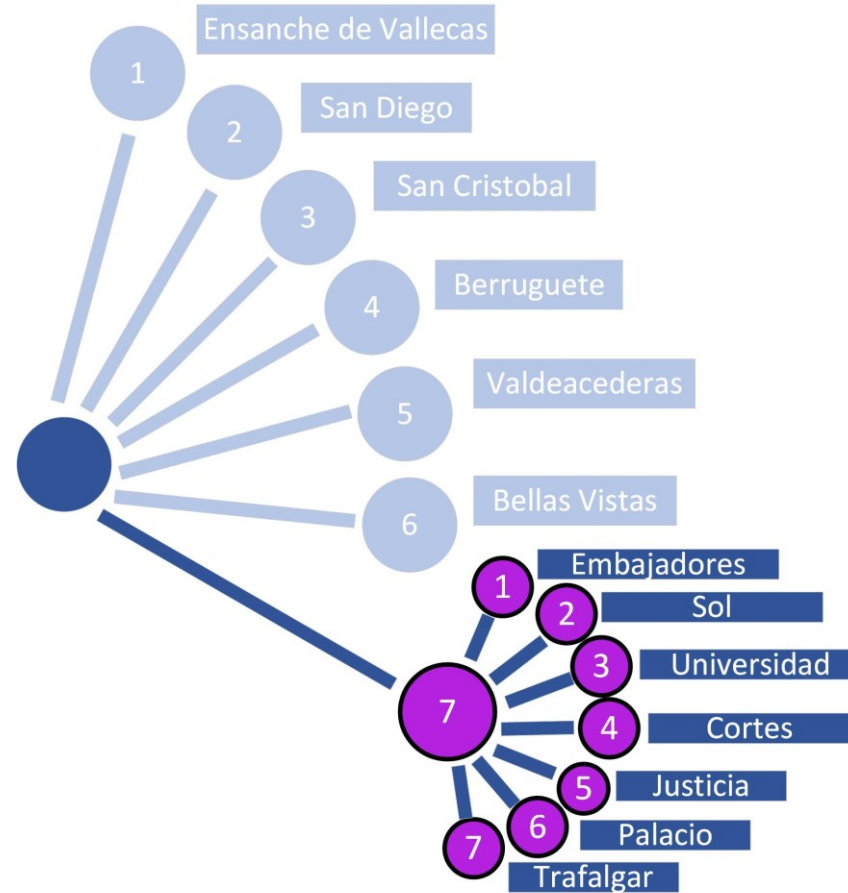
- **The 2nd Most Common Venues are:**

- Theater 2
- Art Gallery 2
- School 1
- Plaza 1
- Student Center 1

Discussion.

The cluster number 7, the purple one, includes way more propitious categories than the other clusters analyzed so the subsequent steps performed in the study will start from this point.

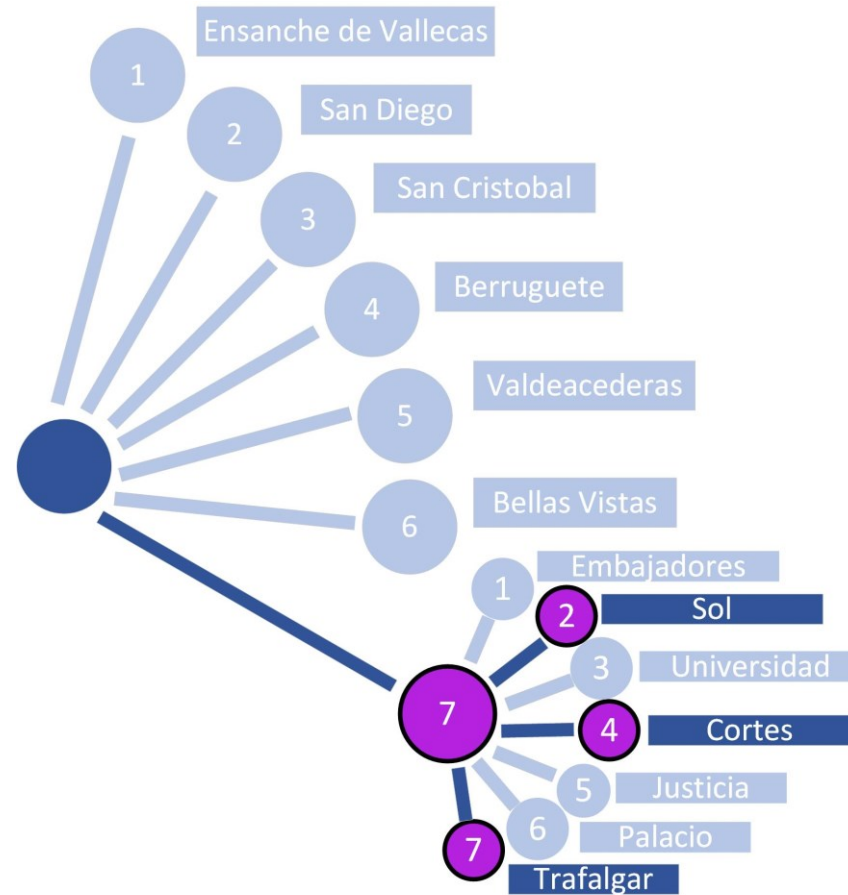
Discussion.



Discussion.

- The borough of Centro is represented by all of its neighborhoods, what could be certainly predictable considering that it concentrates most the cultural and monumental area of the town.
- Although someone could think that maybe an unexpected guest has sneaked into the study, the reality reveals that the borough of Chamberí is a zone with a wide cultural presence, with museums and theaters that increase the social life and popularity of the zone..
- Once the machine learning algorithm has concluded and the best cluster has been selected, the next task is to detect the presence of public libraries in the proximities of the remaining centroids.
- The neighborhoods of Embajadores, Universidad, Justicia and Palacio, all in the borough of Centro, have a public library within their margins and so they should be dismissed.

Discussion.



Discussion.

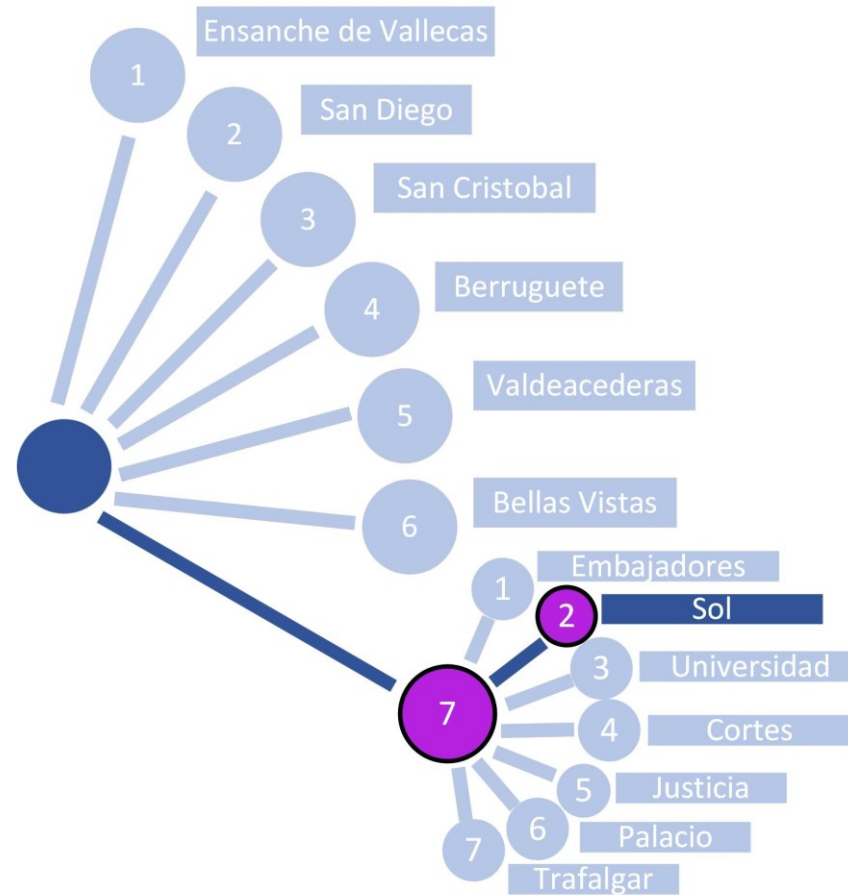
- Finally the mean price of the land in each territory should be evaluated, thereby optimizing the initial minimum investment.

Centro	4.752 €/m2	Chamberí	5.306 €/m2
Chueca-Justicia	5.949 €/m2	Almagro	6.443 €/m2
Huertas-Cortes	5.167 €/m2	Arapiles	4.882 €/m2
Lavapiés-Embajadores	4.226 €/m2	Gaztambide	4.682 €/m2
Malasaña-Universidad	4.910 €/m2	Nuevos Ministerios-Ríos Rosas	4.996 €/m2
Palacio	4.643 €/m2	Trafalgar	5.391 €/m2
Sol	4.605 €/m2	Vallehermoso	5.010 €/m2

Conclusion.

- Beginning the study with a limited set of elements may increment the precision of the prediction.
- By the use of the machine learning algorithm of k-means it has been possible to determine which neighborhoods could fit appropriately to the requirements initially established.
- In this case, teaching and culture are the key of success, neighborhoods with these elements will improve the viability of the business.
- Finally, it is necessary to bear in mind the economical aspect of the operation, and the initial investment, in this case, the cost of the commercial rental of the store.
- At the end of the journey there is only a neighborhood that meets the expectations proposed at the beginning of the project.

Conclusion.



Conclusion.

- The neighborhood of **Sol** is integrated in the borough of **Centro**. With a population of 7.622 citizens is the most centric neighborhood of the capital and the reference of all the peninsular roads (the kilometer zero). Was designed in its actual shape in the middle of the 19th century.

References.

- [Foursquare Official Site](#)
- ["Madrid, con 497, adelanta a Londres y París en número de librerías, sector sin bajas en la capital desde 2015" \(Europa Press\)](#)
- ["Así es el plan de rescate de las librerías de Madrid" \(El Mundo\)](#)
- ["Hábitos de Lectura y Compra de Libros en España | 2020" \(Federación de Gremios de Editores de España\)](#)
- [Madrid City Council Official Site](#)
- ["Distrito Centro" \(Ayuntamiento de Madrid\)](#)
- ["The neighborhood of Embajadores \(Centro\)" \(Wikipedia\)](#)
- ["The neighborhood of Sol \(Centro\)" \(Wikipedia\)](#)

References.

- [The neighborhood of Universidad \(Centro\)" \(Wikipedia\)](#)
- [The neighborhood of Cortes \(Centro\)" \(Wikipedia\)](#)
- [The neighborhood of Justicia \(Centro\)" \(Wikipedia\)](#)
- [The neighborhood of Palacio \(Centro\)" \(Wikipedia\)](#)
- [The neighborhood of Trafalgar \(Chamberí\)" \(Wikipedia\)](#)
- ["Land price in Centro borough" \(Idealista.com\)](#)
- ["Land price in Chamberí borough" \(Idealista.com\)](#)
- ["Retail spaces for rent in Sol" \(BELBEX\)](#)

The scent and the inspiration.

- Thanks to Murtaza Haider author of the book "Getting Started with Data Science. Making Sense of Data with Analytics" and to Pearson plc and IBM Press, its publishers, for bringing to my mind me the germ of the thematic of the project.
- Thanks to Maison Crivelli and its beautiful creation "Papyrus Moléculaire" which has been a pleasure for my pituitary these days of work.

