

València's neighborhoods: a slightly analytical approach.

An analysis of kinds of neighborhoods in València and the suitability of setting a restaurant.

*Created by **David Moreno Valls**, MSc Industrial Engineer student from UPV.*



Table of Contents

1. Introduction and welcome
2. Data acquisition and cleaning
3. Methodology
4. Analysis
5. Results and Discussion
6. Conclusion
7. Acknowledgements

Introduction and welcome

Hello and welcome to my Data Science's Capstone Project!

In this project you will see my report about the topic that I've chosen for València. This report will include a neighborhood analysis, which is about the 10th most common venues categories of each neighborhood, also the conclusion of which restaurant topic has been chosen, by setting the dataset for having graphs and data frames to make the arguments stronger discussed in this notebook. Hope you enjoy reading this document.

Background

València, a sunny city placed in the Peninsula Ibérica's east coast, is one of the best cities to live in Europe, as some reports from important brands have declared, i.e. this article from globexs.com. In order to get in place, bunches of restaurants have already settled in Valencia. In addition, local franchise restaurants are raising and becoming very popular in the city and all the region. Although it's not the most populated spanish city (it's the third), the city receives nearly 2 million tourists each year, read this article from [Levante](#) (in spanish) - and this numbers have just started to grow up! -.

Therefore, there are plenty of reasons to open a business in Valencia. On the one hand, you have a 1.5 million inhabitants city so locals will give you an opportunity to grow up. On the other hand, you have the high amount of tourists coming to the city and bringing you the chance to serve them the best quality service that you can.

Problem

Suppose some investor wants to open a restaurant in València. Capital is available but there's no preference for choosing a topic (i.e. mexican, japanese, tapas, greek). Instead of giving importance to which topic would be chosen, stakeholders are more interested in reporting an analytic of which category will fit the most in each neighborhood - but where? How? Which restaurant will suit better?

These are the key questions about the problem.

Data acquisition and cleaning

Data sources

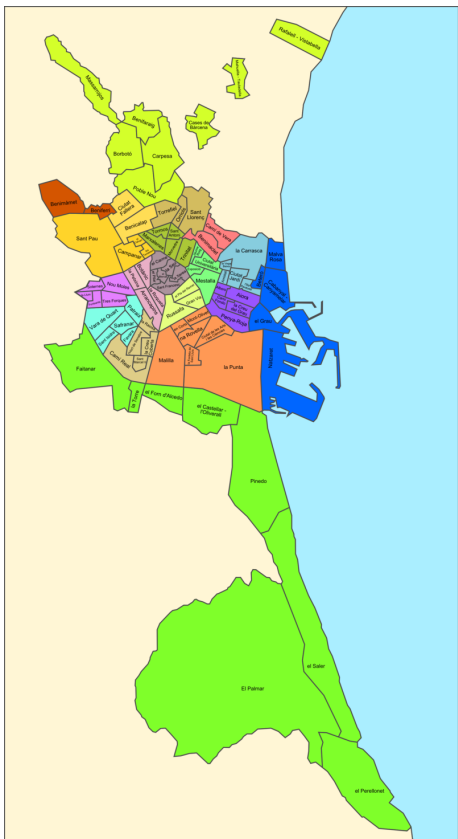
There is a couple of sources that are going to be used. At the first part, for geolocalization purposes, a Wikipedia page would be chosen, where the different districts and neighborhoods are posted. At the second part, google API will be requested for get the latitude and longitude coordinates from each neighborhood. In the last part, the foursquare API, as is a good source of venue data, will become the main source of statistics about the neighborhood's features.

You can find the wikipedia page from [here](#). Other sources of data were taken by an API request, so there is not published in a web.

Data cleaning and processing

First of all, data from the wikipedia's table needs to be cleaned. The original table's source is as it's shown in the next figure.

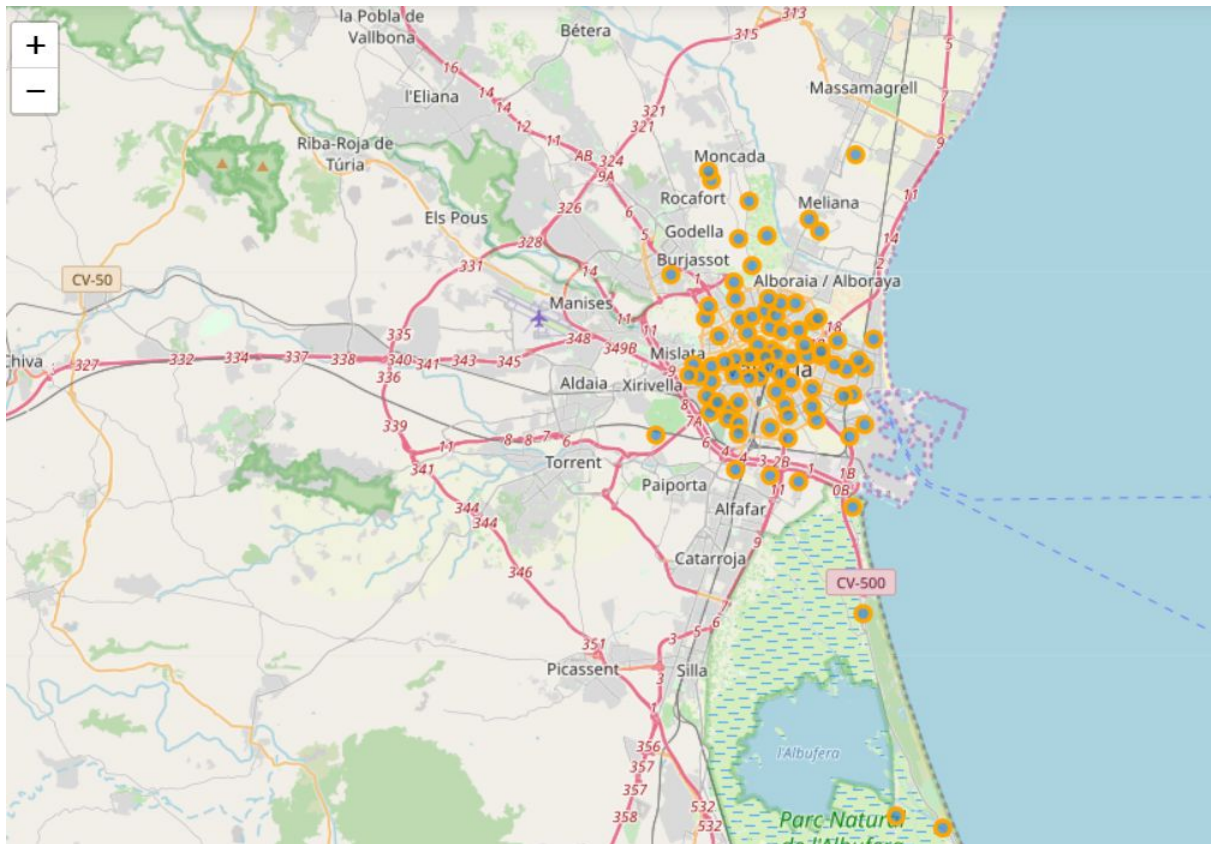
Núm.	Districte	Extensió (ha)	Habitants	Densitat (hab/km²)	Barris
1	Ciutat Vella	169,0	26.769	15.841	la Seu; la Xerea; el Carme; el Pilar; el Mercat; i Sant Francesc
2	l'Eixample	173,3	42.406	24.468	Russafa; el Pla del Remel; i Gran Via
3	Extramurs	197,2	48.287	24.491	el Botànic; la Roqueta; la Petxina; i Arrancapins
4	Campanar	531,9	37.562	7.062	Campanar; les Tendetes; el Calvari; i Sant Pau.
5	la Saïdia	194,4	46.818	24.084	Marxalenes; Morvedre; Trinitat; Tormos; i Sant Antoni
6	el Pla del Real	169,3	30.116	17.792	Exposició; Mestalla; Jaume Roig; i Ciutat Universitària.
7	l'Olivereta	200,9	48.221	24.001	Nou Moles; Soternes; Tres Forques; la Font Santa; i la Llum.
8	Patraix	289,1	57.573	19.914	Patraix; Sant Isidre; Vara de Quart; Safranar; i Favara
9	Jesús	298,5	52.261	17.509	la Raiosa; l'Hort de Senabre; la Creu Coberta; Sant Marcel·lí; i Camí Real
10	Quatre Carreres	1.132,6	73.664	6.504	Montolivet; En Corts; Malilla; Font de Sant Lluís; Na Rovella; la Punta; i Ciutat de les Arts i les Ciències
11	Poblat Marítims	396,8	57.590	14.512	el Grau; el Cabanyal-Canyamelar; la Malva-rosa; Beteró; i Nazaret
12	Camins al Grau	236,7	65.119	27.506	Aiora; Albors; la Creu del Grau; Camí Fondo; i Penya-roja
13	Algirós	297,8	37.191	12.489	l'Illa Perduda; Ciutat Jardí; l'Amistat; la Bega Baixa; i la Carrasca
14	Benimaclet	164,3	29.096	17.712	Benimaclet; i Camí de Vera
15	Rascanya	262,7	52.749	20.079	els Orriols; Torrefiel; i Sant Llorenç
16	Benicalap	222,1	45.460	20.470	Benicalap; i Ciutat Fallera
17	Pobles del Nord	1.519,6	6.529	430	Benifaraig; Poble Nou; Carpesa; Cases de Bàrcena; Mahuella; Massarrojos; i Borbotó
18	Pobles de l'Oest	200,4	13.993	6.983	Benimàmet; i Beniferri
19	Pobles del Sud	3.226,1	20.228	627	el Forn d'Alcedo; Castellar-Oliveral; Pinedo; el Saler; el Palmar; el Perellonet; la Torre; i Faltanar



By setting up the data, the target is to have a single-neighborhood row table with their latitude and longitude coordinates, in order to fit the data into a map. There were some problems while processing, as the geolocator couldn't find all the adresses at once, a couple of attempts were tried with the non-found locations and then in a last chance locations were written manually. The table result is as shown.

Districte	Barris	latitude	longitud
Ciutat Vella	la Seu	39.476749	-0.375003
Ciutat Vella	la Xerea	39.4746544	-0.3721794
Ciutat Vella	el Carme	39.4782212	-0.3810013
Ciutat Vella	el Pilar	39.4739726	-0.382343
Ciutat Vella	el Mercat	39.4738165	-0.3778455

Finally, the table is cleaned and the markers were in the map correctly, as shown in the next map.



Methodology

In this project, we'll be focused on detecting different types of areas in Valencia depending on which kind of venues are settled. We'll be particularly interested in those areas where its number of venues and emblematic places are higher than in other areas with less venue density. This is because the main clients will be both tourists and locals.

Firstly, the required **data** is collected: **location and type (category) of every venue within 250 m from each neighborhood**.

Secondly, in the analysis will be to calculating and exploring '**clusters**' across the different neighborhoods of València to get different comparison from each neighborhood.

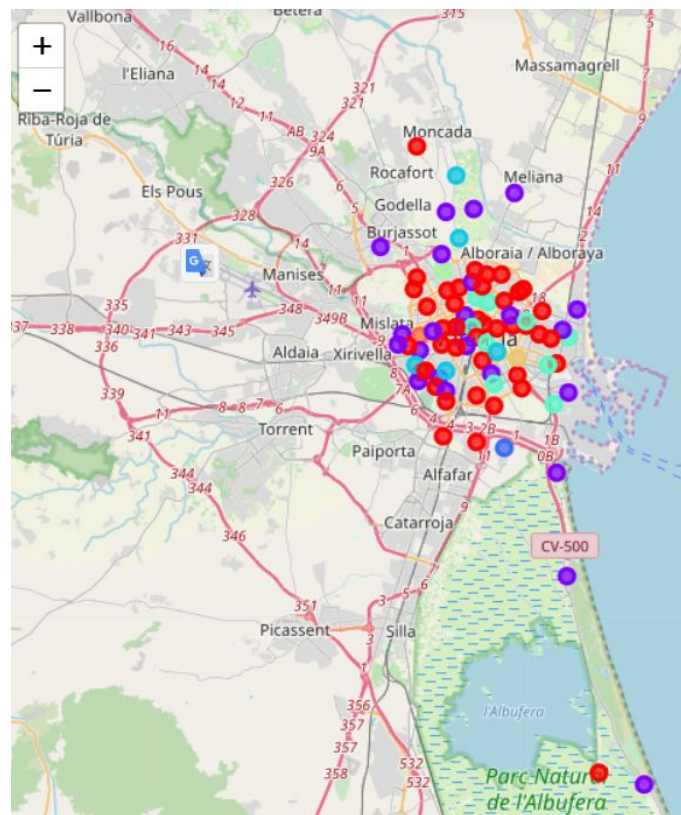
Finally, we will focus on the most promising areas and within those create **bar plots** established on discussion with stakeholders.

Analysis

Let's perform some basic understandable data analysis and derive some additional info from our raw data. First let's see an example of the most common venues in every neighborhood:

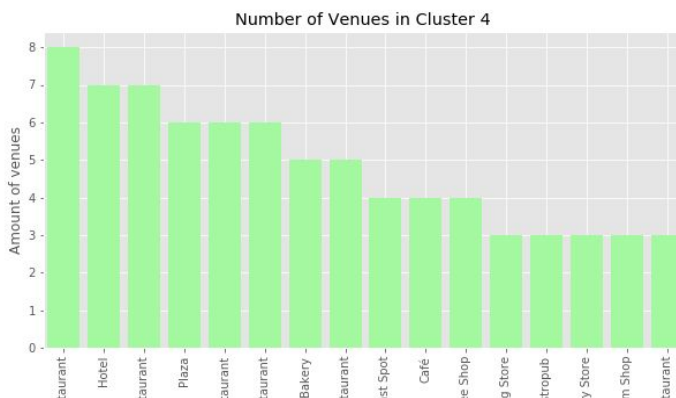
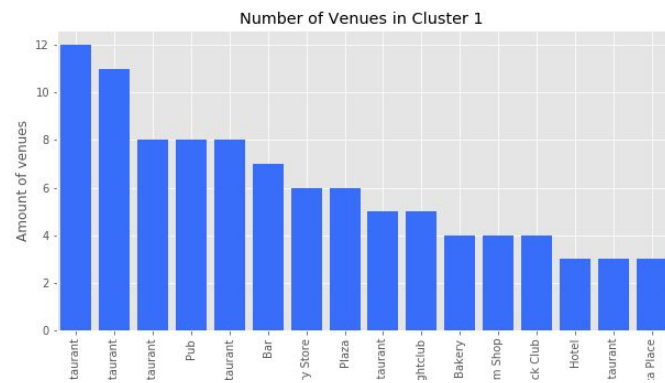
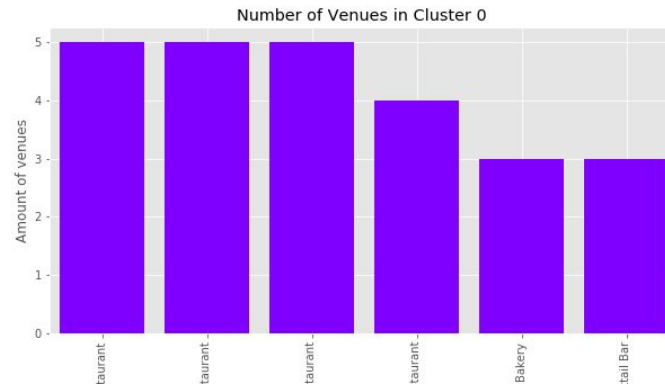
n	Districte	Barris	latitude	longitude	Cluste	1st Most Common Venue	2nd Most Common Venue
0	Ciutat Vella	la Seu	39.476749	-0.375003	0	Spanish Restaurant	Restaurant
1	Ciutat Vella	la Xerea	39.4746544	-0.3721794	0	Plaza	Spanish Restaurant
2	Ciutat Vella	el Carme	39.4782212	-0.3810013	1	Mediterranean Restaurant	Tapas Restaurant
3	Ciutat Vella	el Pilar	39.4739726	-0.382343	1	Coffee Shop	Italian Restaurant
4	Ciutat Vella	el Mercat	39.4738165	-0.3778455	4	Tapas Restaurant	Plaza
5	Ciutat Vella	Sant Francesc	39.4701823	-0.3755728	0	Restaurant	Hotel

Once the venue data is given, neighborhoods were clustered into different groups, depending on the venue affinity. This process is done by a **K-means algorythm**, given the unlabelled weights of each venue in each neighborhood. The result is as shown in the map below:



Clusters that we have to take are those which their neighborhoods have the most emblematic places, so more tourist are attracted, and discard the "unique clusters", which have up to 3 neighborhoods.

In order to get an approach of which neighborhoods have more emblematic places, **City center** and the **Harbour** will be on top of tourism attraction. It corresponds to the boroughs of *Ciutat Vella*, *Camins al grau* and *Poblats Marítims*. The clusters which includes neighborhoods from these boroughs are **Cluster 0, 1 and 4**. Let's get a deeper analyse from them.



Results and Discussion

Our analysis shows that although there is a great number of restaurants in Valencia, some different clusters determine that not all types of restaurants are equally distributed. Instead of that, Clusters that we'd analyze later seems that besides the popularity of spanish cuisine and paella restaurants, there are a bunch of italian restaurants, also pubs and nightclubs. In this case we see that the target neighborhoods were optimal chosen by the clusters that we've selected.

Cluster number 0, as it is the most popular, seems like it has a gap of seafood restaurants, which it's far from other kind of more popular restaurants. Other kinds of restaurants such as afghanistan restaurants are discarded due to their global non-popularity in the city (there are not usual).

Cluster number 1 is similar to the last cluster, but it's in the lower frequent venues where it marks the difference. In this case, pubs and nightclubs are more common places. Is in this cluster where it is recommended to set up a young-thematic restaurant, which includes discounts for a night club, etc.

Finally, cluster number 4 belongs to a more familiar kind of neighborhoods, where high-income families are living, as we can appreciate because of the amount of gastropubs. In this case, a more elegant restaurant is recommended.

Thanks to the report made, where among the different kind of neighbourhoods, is recommended to settle a restaurant and how it is needed to be focused based on which cluster it belongs is concluded.

Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually end up in a location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Valencia's similar areas in order to aid stakeholders in narrowing down the search for optimal locations for a new restaurant without any thematic already defined. By calculating venues density distribution from Foursquare data, we have first identified general clusters that justify further analysis, and then generated extensive collection of locations which satisfies some basic requirements regarding existing nearby restaurants.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into account additional factors such as attractiveness of each location (proximity to parks or malls), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.

Acknowledgements

I would like to thank all that I learnt to **Coursera** and specially to the teachers that taught me how to think methodology but by the data scientist side. As an engineer I'm used to have a bit different perception on how to manage data, and this course bring me another point of view of how are data managed.