

Capstone Project - The Battle of Neighborhoods: Opening a new pharmacy in the Seville, Spain metropolitan area

Fernando Delgado Chaves

1. Introduction

The present project is the final assignment of the Capstone, the very last course of the series of 9 courses composing the IBM Data Science Professional Certificate.

1.1 Description of the problem and a discussion of the background

The primary objective of the pharmaceutical industry is to research, discover and develop new medicines that can improve the health and quality of life of people who suffer from a health problem. But, at the same time, the sector has become one of the main drivers of the economy in countries with established pharmaceutical companies, such as Spain, where it is the industrial sector that invests the most in R&D, being responsible for one out of every five euros spent in this strategic area for any developed economy.

A key factor is that the pharmaceutical sector is distinguished by its tractive effects on other sectors of the economy, something that is especially revealing in the area of employment generation. Thus, the pharmaceutical industry generates direct jobs all around the globe, including professionals with university studies, and is the high technology sector with the most employment.

1.2 Business problem

Opening a new pharmacy, also known as chemist or drugstore, is a high-risk investment that involves strategic decision making. New stores are required to satisfy the needs of communities with not so many facilities available. Usually, pharmacies are positioned at even distances so they become accessible to a large number of individuals, according to a population criterion.

In addition, it is to be considered that whether a pharmacy is close to a Doctor's office or hospital it will be more likely to be visited by patients after their medical appointment. Analogously, gyms or sport centers often recommend food supplies and other fitness protein which are provided in pharmacies.

For the sake of simplicity, the following aspects will be taken into consideration for selecting an appropriate place to open a new pharmacy:

- Demand for pharmacies and medicinal products
- Closeness to hospitals or medical offices

- Closeness to gyms or sport centers

1.3 Target audience

The main target audience for this project will be developers and investors in the Seville, Spain metropolitan area. It is useful for chemists and pharmaceutical companies who want to open a new pharmacy in Seville, as well as for drug providers who are interested in new sales.

1.4 Description of the data and how it will be used to solve the problem

In this project geospatial information of the metropolitan area of Seville, Spain will be selected in order to open a new pharmacy. This data was retrieved from the Foursquare API, which includes the location of an area and its facilities, including the location of other pharmacies, hospitals, medical offices and other places of interest. The Python Folium package was used to plot such information on a map highlighting the facilities of each area.

Based on the features of each area, the aim of this project is to determine a ranking of the best areas or neighborhoods to start a new pharmaceutical business.

2. Methodology

First, geospatial information was retrieved for the spots of interest. The list of districts and neighborhoods of Seville were obtained from: <https://urbansevilla.es/distritos-y-barrios-de-sevilla/> to be used as input for a geolocalization service. This list was submitted to *Geocode* in order to obtain the information on their latitude and longitude. Additionally, the Foursquare API was used to retrieve the list of hospitals in the city and other venues close to the different neighborhoods.

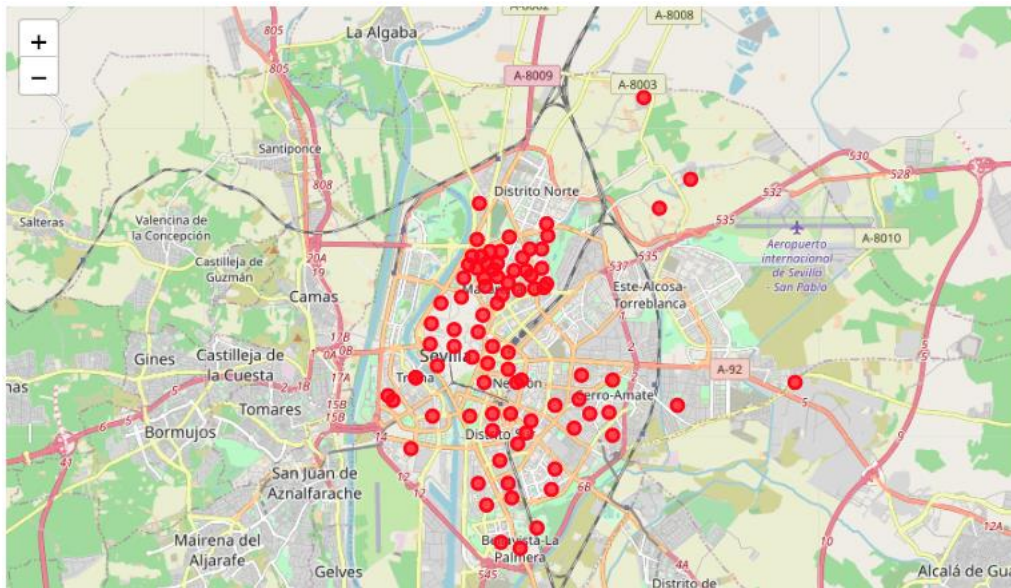
Second, the different neighborhoods were analyzed and finally clustered in order to obtain the neighborhoods that present a low number of pharmacies but at the same time contain a large number of hospitals and gyms.

3. Results

A sample of the list of Sevilla neighborhoods to be taken into consideration is shown below. It is composed of 85 neighborhoods grouped in a total of 10 districts. The geographical coordinates of these neighborhoods was retrieved using *Geocode*.

	District	Neighborhood	Address	Longitude	Latitude
0	Bellavista – La Palmera	Bellavista	Bellavista, Bellavista – La Palmera, Seville	-5.974072	37.344722
1	Bellavista – La Palmera	Heliópolis	Heliópolis, Bellavista – La Palmera, Seville	-5.983839	37.354551
2	Bellavista – La Palmera	Elcano-Los Bermejales	Elcano-Los Bermejales, Bellavista – La Palmera...	-5.979620	37.346070
3	Bellavista – La Palmera	Sector Sur-La Palmera-Reina Mercedes	Sector Sur-La Palmera-Reina Mercedes, Bellavis...	-5.986303	37.359483
4	Bellavista – La Palmera	Pedro Salvador-Las Palmeritas	Pedro Salvador-Las Palmeritas, Bellavista – La...	-5.976616	37.356237

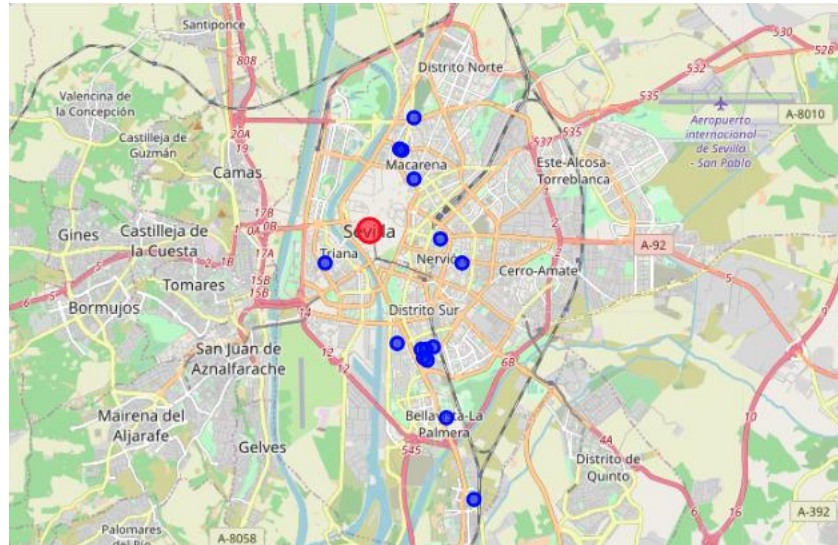
The list of Sevilla neighborhoods to be taken into consideration are geolocated in the map shown below.



The list of Sevilla hospitals to be taken into consideration is shown below. These medical centers were retrieved using the Foursquare API.

- | | |
|----|---|
| 0 | Hospital Infanta Luisa |
| 1 | Hospital Universitario Virgen del Rocío |
| 3 | Hospital Infantil Virgen del Rocío |
| 8 | Hospital de Rehabilitación y Traumatología |
| 12 | Hospital Viamed Santa Ángela de la Cruz |
| 14 | Hospital Victoria Eugenia - Cruz Roja Sevilla |
| 15 | Hospital Santa Isabel |
| 17 | Hospital Regional Virgen Macarena |
| 21 | Hospital San Juan de Dios |
| 22 | Hospital Duques del Infantado |
| 23 | Hospital Sagrado Corazón |
| 25 | Hospital Provincial de San Lázaro |
| 26 | Hospital Quirón Sagrado Corazón |
| 29 | Hospital Militar - Sevilla |
| 32 | Sala Ecografía Hospital Virgen Macarena |

The list of Sevilla hospitals to be taken into consideration are geolocated in the map shown below. In red it is represented the city center, whereas the hospitals are represented in blue.



A sample of the venues retrieved in the surroundings of each neighborhood is shown below.

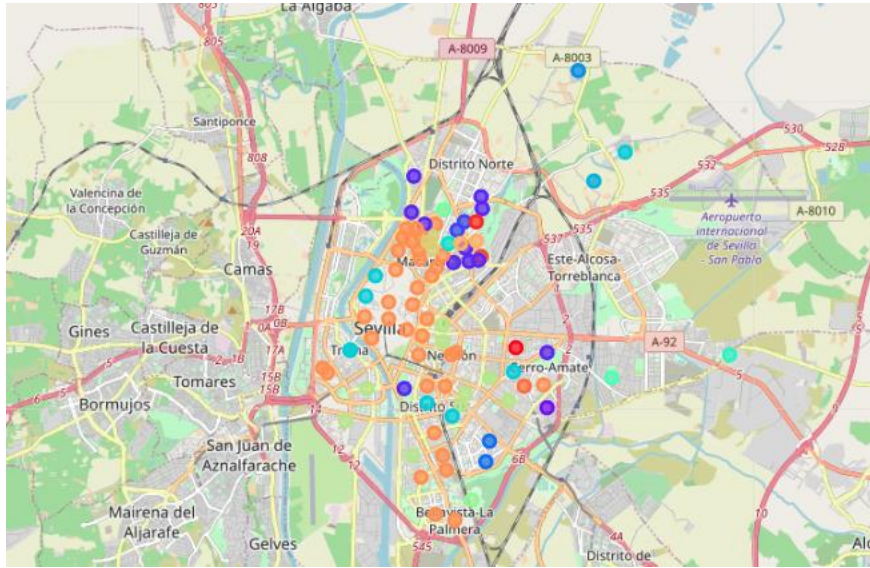
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bellavista	37.344722	-5.974072	Helarte	37.346009	-5.977549	Ice Cream Shop
1	Bellavista	37.344722	-5.974072	Ignacio Vidal	37.343917	-5.977493	Restaurant
2	Bellavista	37.344722	-5.974072	La Tasca del Pintor	37.343918	-5.977893	Tapas Restaurant
3	Bellavista	37.344722	-5.974072	Cervecería Huracán	37.345952	-5.978268	Bar
4	Bellavista	37.344722	-5.974072	Don Matteo (Tapas & Pizzas)	37.343981	-5.976766	Pizza Place

To this list, the previously-retrieved hospitals were added.

Then, neighborhoods and their venues were explored in order to analyze the similarities between them. The top 100 most frequent venue was obtained for each neighborhood. An example is shown below.

	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
0	Aeropuerto Viejo	Garden Center	Gaming Cafe	Furniture / Home Store	Fried Chicken Joint	French Restaurant	Food & Drink Shop	Food	Flea Market	Fish Market
1	Alfalfa	Tapas Restaurant	Plaza	Hotel	Spanish Restaurant	Ice Cream Shop	Clothing Store	Restaurant	Mediterranean Restaurant	Bar
2	Amate	Bar	Chinese Restaurant	Tapas Restaurant	Spanish Restaurant	Steakhouse	Winery	Fast Food Restaurant	French Restaurant	Food & Drink Shop
3	Bami	Hospital	Spanish Restaurant	Tapas Restaurant	Soccer Field	Restaurant	Plaza	Fast Food Restaurant	Food & Drink Shop	Food
4	Barriada de Pineda	Gym	Brewery	Mountain	Golf Course	Fish & Chips Shop	French Restaurant	Food & Drink Shop	Food	Flea Market

In addition, neighbors were clustered to further predict these similarities. A total of 20 clusters were identified, which are shown in different colors in the map below.

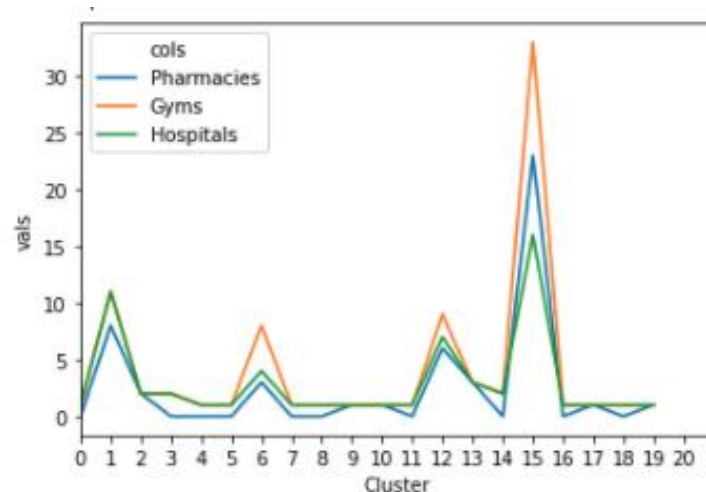


In addition, neighbors were clustered to further predict these similarities. A total of 20 clusters were identified, which are shown in different colors in the map below.

In order to evaluate those clusters where pharmacies are scarce but hospitals and gyms are abundant the difference between the absolute number of each was performed according to the following formula.

$$\text{Difference} = (\text{hospitals} + \text{gyms}) - \text{pharmacies}$$

This is represented by the following line plot. The higher the gap between the pharmacies line (blue) and the gyms and hospitals lines (respectively orange and green), the better the cluster for locating a new pharmacy.



4. Discussion

According to the presented results, the best cluster that could be considered for starting a new pharmaceutical business would be cluster 15. This cluster contains the following neighborhoods:

0	Bellavista
1	Heliópolis
2	Elcano-Los Bermejales
4	Pedro Salvador-Las Palmeritas
6	El Arenal
7	Encarnación-Regina
8	Alfalfa
9	San Bartolomé
11	San Gil
12	Museo
13	Santa Catalina
15	San Julián
18	Juan XXIII
32	Pío XII
34	El Carmen
35	Cruz Roja-Capuchinos
40	Polígono Norte
41	La Paz-Las Golondrinas
42	La Palmilla-Doctor Marañón
44	Macarena 3 Huertas-Macarena 5
46	El Cerezo
47	Doctor Barraquer-Grupo Renfe-Policlínico
48	Retiro Obrero
51	León XIII-Los Naranjos
54	Las Avenidas
55	Nervión
56	San Bernardo
58	La Florida
59	La Buhaira
61	San Roque
88	Huerta de la Salud
90	Giralda Sur
93	Tabladilla-La Estrella
94	Bami
101	Barrio León
102	El Tardón-El Carmen

As shown in the graph, this cluster exhibits the lowest number of pharmacies and at the same time, exhibits a large number of hospitals and gyms whose users require medical attendance. Hence, there is a need for this kind of business in the neighborhoods in this cluster.

A limitation of the present study would be not taking the population rate into consideration. According to the Spanish law, pharmacies should be displayed in agreement with the population of the areas.

5. Conclusion

The neighborhoods shown above, corresponding to cluster 15, are the most appropriate or setting up a new pharmacy.