Capstone Project

The Battle of Neighborhoods

Fidel Navarro Salazar

1. Introduction

1.1 Background and Problem

Mexico City (CDMX) is one of the biggest cities in the world with a total population of 8 918 653 inhabitants and a density of 6 000 / km^2. Due to the high demand of healthcare centers in the city it is very common to find these services saturated, and with the increase of covid-19 cases this situation gets worse by da day.

1.2 Problem

This project's purpose is to help people find and choose healthcare centers based on their location, availability, user experience and necessities in a big and busy city like Mexico City (CDMX).

Large cities such as CDMX where the population is approximately 9 million inhabitants (one of the most inhabited cities in the world) may have traffic and access problems to people in public areas. Furthermore, with the increasing demand of health services, these can easily exceed their attention capacities. For this reason, this project seeks to solve these problems, creating a program that allows viewing of the nearest health centers with adequate services and capacity.

The project will work with a dataset of Hospitals and Health Centers provided by the government from Mexico City and user data from Foursquare.

1.3 Interest

Users, such as local population and tourists would be very interested in getting recommendations of near by hospitals in case of an emergency.

2. Data acquisition and cleaning

2.1 Data sources

The main data of hospitals was taken directly from the database of the city Government. This dataset contains information from 27 mayor public hospitals of the city. Information as name, coordinates (latitude and longitude), person in charge, address and geopoint can be found in the dataset. A mayor inconvenient with the data set provided was the lack of information on private hospitals and healthcare centers. This missing information was complemented using the database of Foursquare. Additionally, Foursquare was used to obtain the user information, such as rating, tips, location, and distance from the user.

2.2 Data cleaning

Both data sets where downloaded and merged by hospital name. Undesired data was removed and was organized by their distance from the user in ascending order.

A mayor problem with the dataset was the lack of ratings provided by users. Another issue was the data provided my Foursquare since it included venues such as veterinary hospitals and cellphone/computer hospitals. This problem was solved by searching and eliminating venues with key words such as dog, cat, animals, veterinary, cellphone, laptop, tablet, and computer in their name.

3. Data visualization

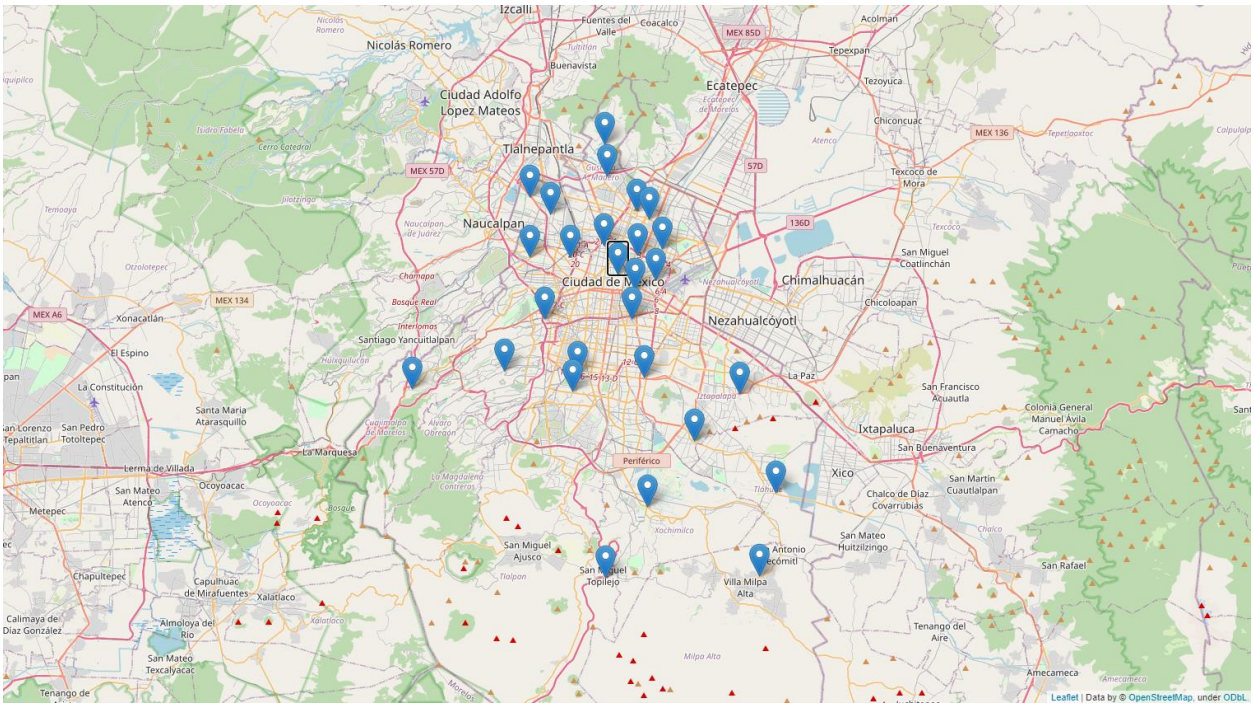

Figure 1: Map with hospitals and healthcare centers in Mexico City (CDMX).

We can visualize in figure 1 most of the hospitals available in CDMX. From this figure we can observe that most of the hospitals and healthcare centers concentrate in the northern part of the city, this could represent a mayor problem when working with users in the southern part. To solve this problem, we can search venues in a bigger radius, and to cut the high amount of venues we can work with the 5 closest hospitals and healthcare centers.

4. Data analysis

Since most hospital do not count with a user rating it is not convenient to recommend a hospital using this parameter. It was decided to use the user's tips and number of tips to make a better comparison between venues. User's ratings were incorporated in later

parts of the code as a secondary parameter, it is only considered when a hospital is close by other healthcare centers with a similar amount of user's tips.

5. Model

When a user's address is acquired the model searches for the five closest hospitals and healthcare centers. Later, the data from the different venues is organized in a data frame by their distance form the user.

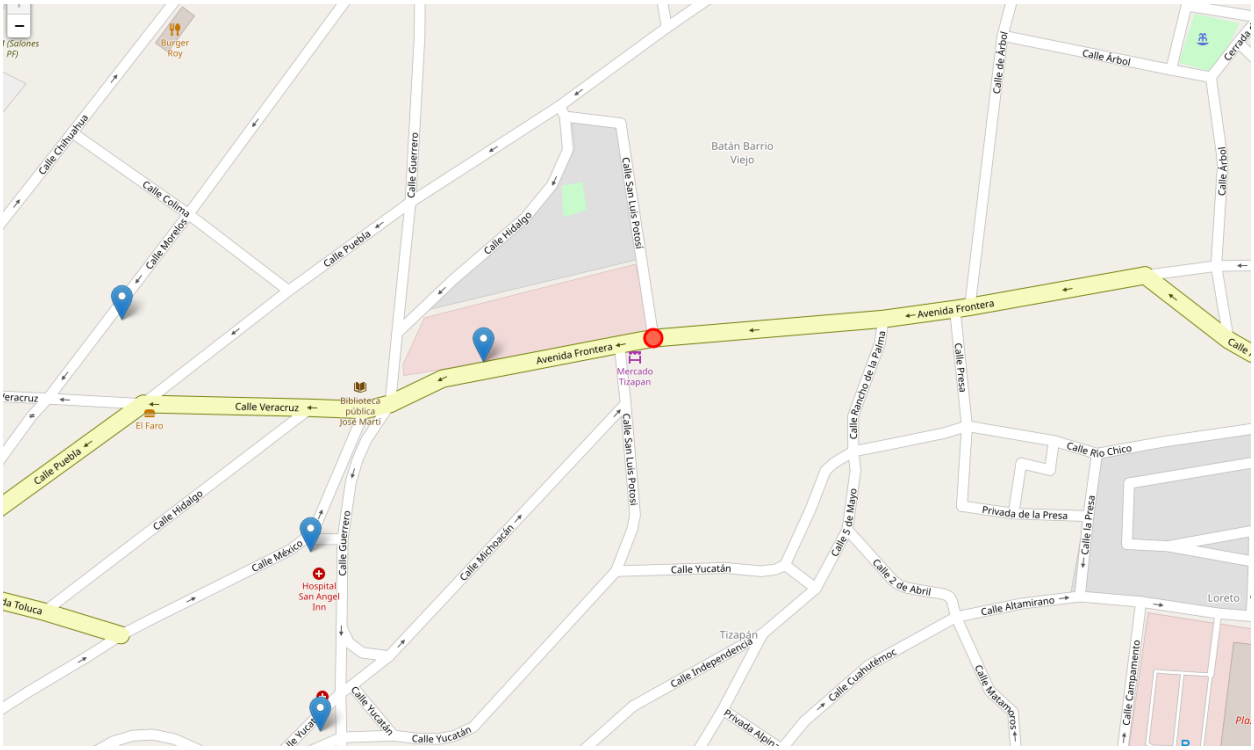

Figure 2: Example of hospitals (blue) near user (red).

In figure 2 we can observe the closest hospitals and healthcare centers from a user. Once this information is organized in a data frame, we obtain the number of tips and ratings from Foursquare. Since most hospitals do not contain a rating, we start comparing the number of tips from each venue and we organize once again the data frame by number of tips. If a venue contains a rating superior to 8.0 and a number of tips greater than the average of tips from the venues found, we recommend that

hospital or healthcare center to our user. If neither of the venues has both conditions, we recommend a venue based on which has the greatest number of tips.

6. Solution

Once our model was coded using Python several tests were conducted to see if the code was working as intended.

First, the user would type its address, as shown in the example of figure 3.

```
address = 'Calle Norte 1, Isidro Fabela, Pedregal de Tepepan, Tlalpan, Ciudad de México, 14030, México'
```

Figure 3: Address of Mexico City.

The model will work on its own collecting and merging data from Foursquare, and the recommendation would be shown by the program as shown in figure 4.

```
The recommended hospital is: Hospital Sedna
Address: Hospital Sedna, Lateral Periférico, Bosques de Tetlameya, Coyoacán, Ciudad
de México, 14388, México
Tip: En general el hospital es buenísimo, la atención, las enfermeras todo me encant
ó ahora que nació mi bebé, incluso la cesárea no me dolió nada! Mis respetos al dr.
```

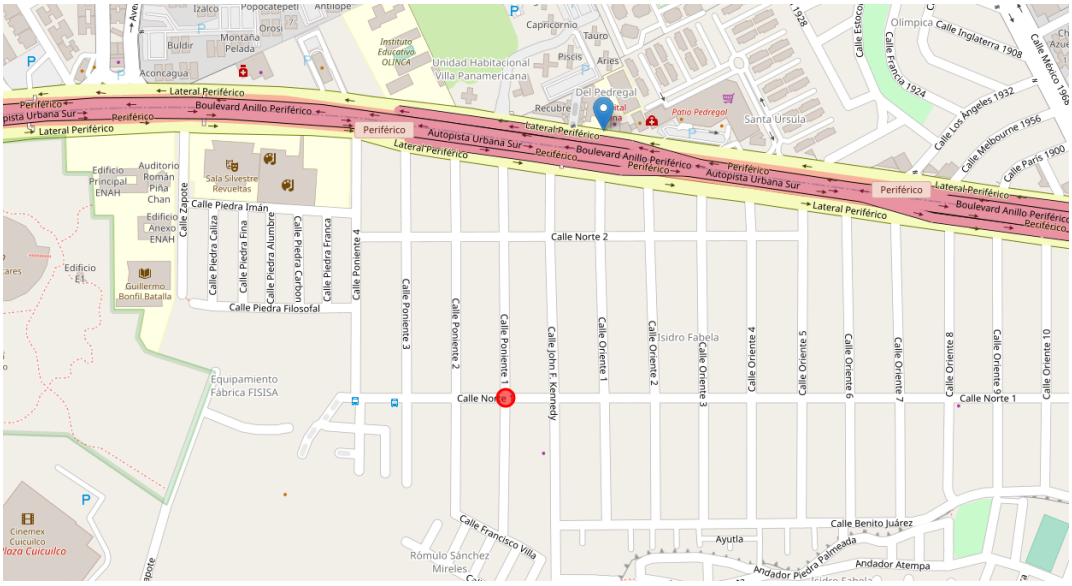

Figure 4: Example of venue recommendation

As seen in figure 4 the information provided to the user contains the name of the hospital or healthcare center, address, most recent tip, and a map indicating the user's location and the venue's location.