Capstone Project
The Battle of Neighborhoods

Fidel Navarro Salazar

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.

Problem

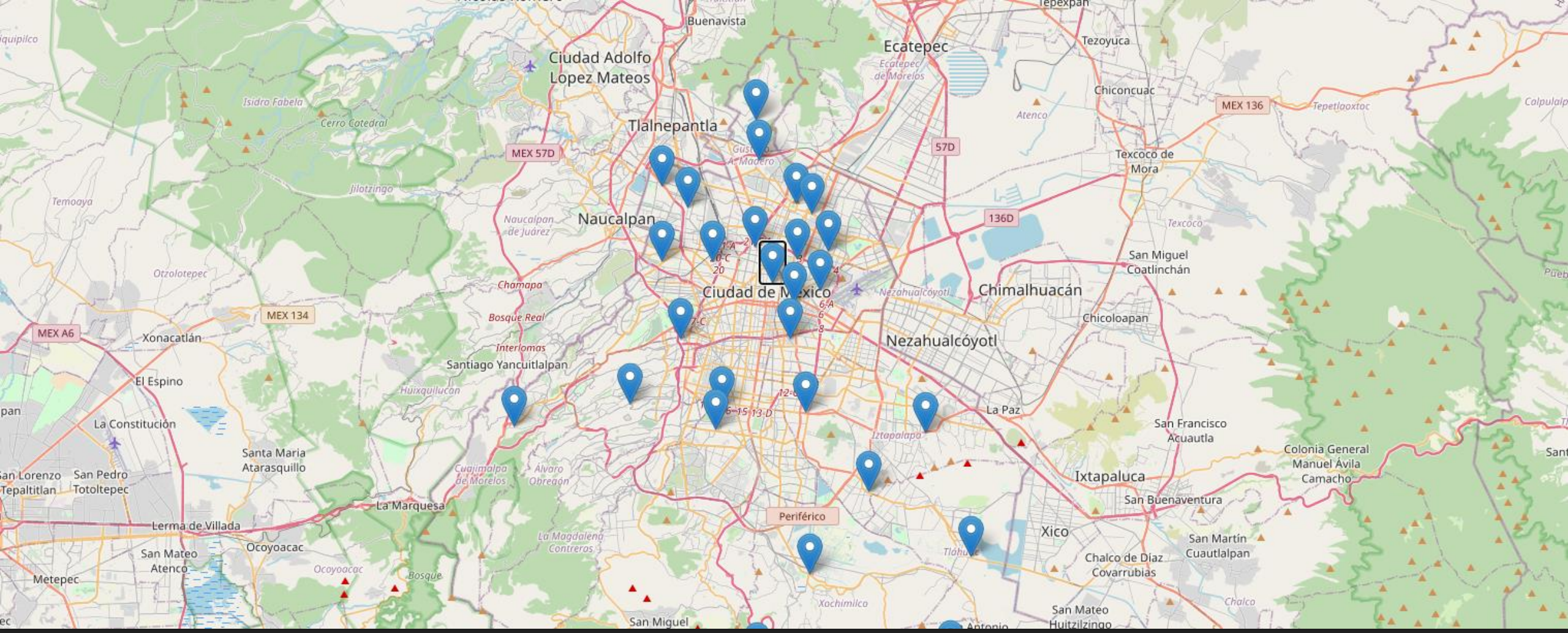
- This project's purpose is to help people find and choose healthcare centers based on their location, availability, user experience and necessities in a big and busy city like Mexico City (CDMX).

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.

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.



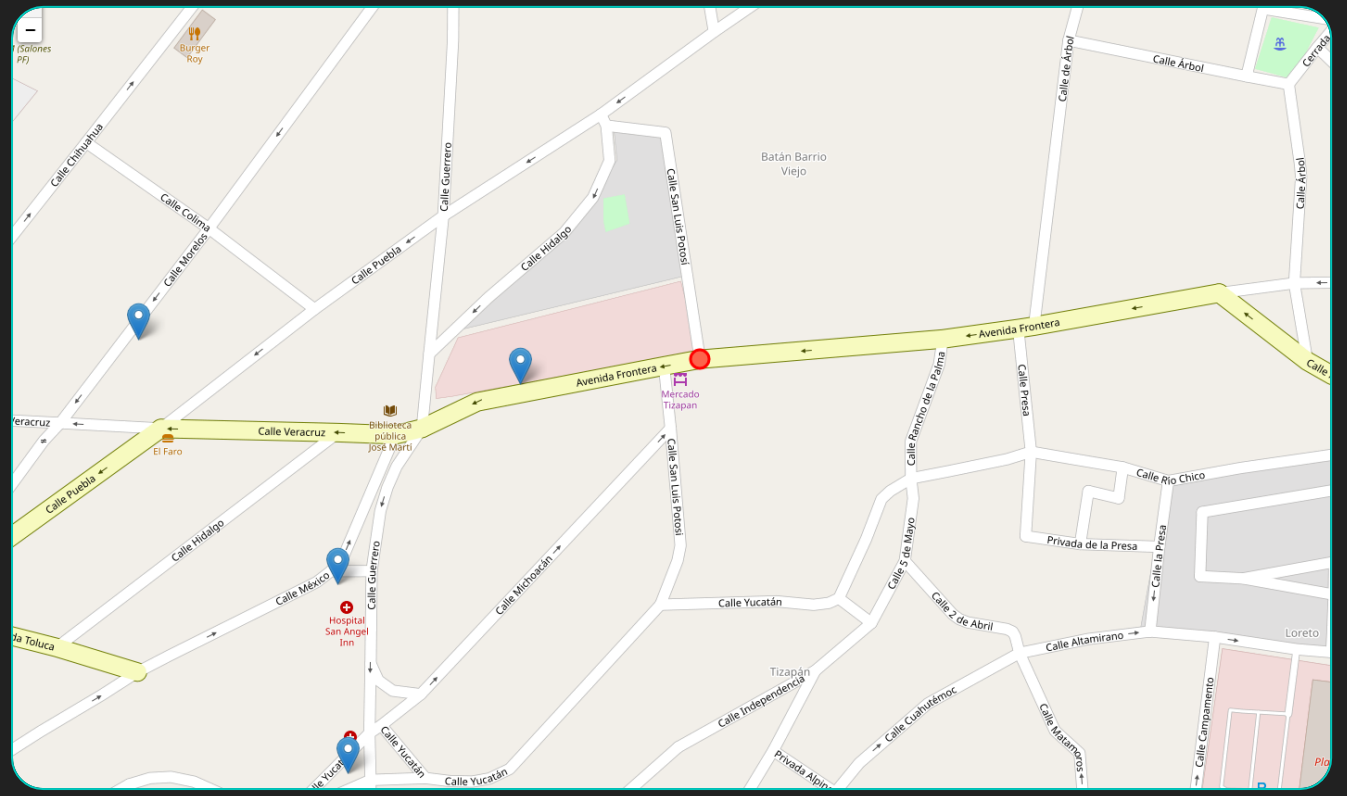
Map with hospitals and healthcare centers in Mexico City (CDMX).

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.

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.



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.

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

- 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.

```
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.
```