Analysis of the comorbidities associated with the COVID-19 pandemic in Mexico and hospital availability in CDMX

José Frías

July 14, 2020

1 Introduction

1.1 Background

The COVID-19 pandemic, caused by SARS-CoV-2 virus, has affected the lives of many people around the world. Governments, organizations and general population are concerned about the evolution of the pandemic and any other information that may be useful in reducing the affectations in their personal and community lives. Governments around the world have taken different preventive and corrective measures to reduce the impact of the pandemic in their societies according to their particular conditions. Among the wide variety of studies concerned with the COVID-19 pandemic, those which analyze the comorbidities associated to the main disease are very useful in protecting vulnerable populations. Another important matter to take into account in order to improve the COVID-19 patients care is the hospital capacity of any city or region under consideration.

1.2 Description of the problem

In the present project we propose the use of public data to analyze two important aspects of the COVID-19 pandemic in Mexico: first, the analysis of the main comorbidities associated to COVID-19 in all the country and, secondly, an analysis of the hospital infrastructure available in Mexico City (CDMX). At the time this report was written, the pandemic in Mexico was in a critical expansion phase and the evolution of the pandemic in the weeks to come is uncertain as can be seen in Figure 1.

With data on the main comorbidities related to COVID-19 disease in Mexico, we aim to use clustering techniques to bunch the 32 states in the country into groups characterized by the affectations of the different comorbidities among their populations. This analysis may be very useful for the federal and local governments to take preventive policies to reduce the risk of health complications among vulnerable populations. The information we will obtain could also be useful for individuals with one or more comorbidities to take their own precautions.

In the second part of this analysis, we aim to use public data and Foursquare to analyze the hospital infrastructure in the 16 municipalities in CDMX. The distribution of the health institutions in Mexico City is not homogeneous and depends on socioeconomic variables. The same is true for the COVID-19 pandemic spread. We aim to compare the total number of COVID-19-positive cases in each municipality in Mexico City with the estimated number of hospitals reported by Foursquare in a representative sector of this municipality. The results of this analysis may be of interest to the city government for the development of strategies in the care of the infected population.

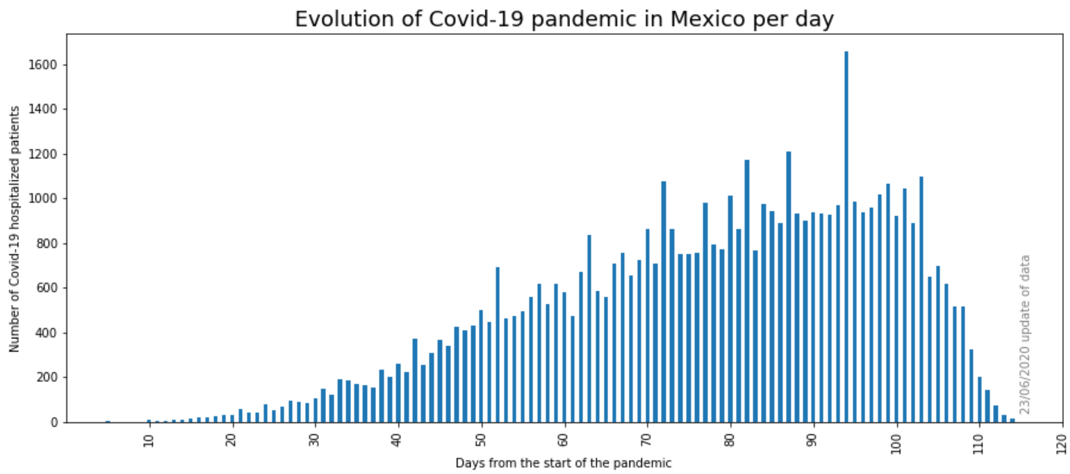

Figure 1: Evolution of COVID-19 pandemic in Mexico

2 Description and preparation of data

2.1 Data acquisition

The first part of the project is based on a dataset named *'casos-asociados-a-covid-19.csv'* that can be found at *https://www.gob.mx/salud/documentos/datos-abiertos-152127* and was updated on June 23, 2020. This dataset contains 384,283 records of persons tested for COVID-19 in Mexico and includes 40 fields containing relevant information for the pandemic statistics. In our analysis we focus on the comorbidities for COVID-19, which constitute nine fields in the dataset, and the final status of each tested person (positive test, hospitalized and dead). Our aim is to use this dataset to analyze the prevalence of the comorbilities among the populations of the Mexican states and cluster them according to this criteria.

For the second part of the project we intend to compare the number of positive cases for COVID-19 in the municipalities of CDMX to the population and the number of hospitals by municipality. The data of coronavirus disease in Mexico City that we will analyze can be found at *https://datos.cdmx.gob.mx/* and was updated on July 1, 2020. This dataset contains 155,722 records of persons tested for COVID-19 in Mexico City and several fields, among which we shall focus on the municipality of origin and status of the records. We also get demographic data of Mexico City municipalities from *https://en.wikipedia.org/wiki/Municipalities_of_Mexico_City*. Finally, we shall use Foursquare to find hospitals and medical clinics in a radius of 2 km around the center of each municipality using the search parameter *categoryId = '4bf58dd8d48988d196941735'*, which corresponds to the *"hospital"* venue category.

For both parts of this project we aim to use the *geocoder* library to find the geographical coordinates of the Mexican states and the municipalities in Mexico City to visualize the results we obtain in our analysis by using the *folium* library.

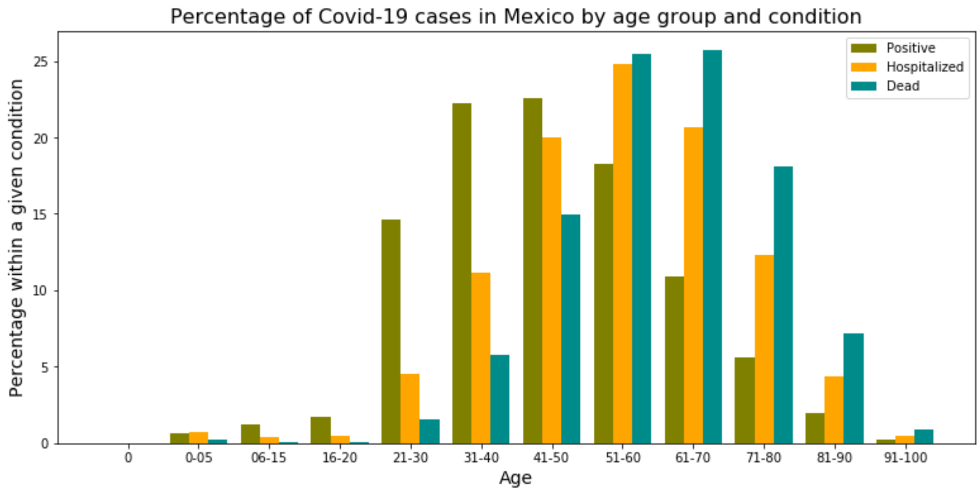

Figure 2: Affectations of the pandemic by age

2.2 Data cleaning

Two of the datasets described in the previous section are huge with hundreds of thousands of records and multiple fields. In any case, we were only interested in a subset of data, therefore, a cleaning data work was carried out.

The dataset on coronavirus tests in Mexico contained 384,283 records, from which we selected those with positive test for COVID-19 and reduce the dataset to 171,371 records. From the 40 fields in the original database we restricted to those of interest to our analysis: state of origin, date of first symptoms, sex, age, comorbidities (diabetes, chronic obstructive pulmonary disease [COPD], asthma, immunosuppression, hypertension, cardiovascular disease [CVD], obesity, chronic kidney disease [CKD] and smoking), and the reported status (positive, hospitalized and dead). The comorbidities fields contained three possible string values 'yes', 'no' and 'unknown', which were converted into floats 1.0, 0.0 and 0.5, respectively. Since the 'unknown' values were a few but they are still part of the records, we decided to assign the value of 0.5 to them. Finally, the reported status field had values NaN, that were corrected to value 0.0. We used the library *datetime* to compute and add a new field with the number of days of the each record since the first case was registered and have notion of the pandemic evolution per day since it started.

From the database containing the 155,722 coronavirus-test results of people from the Valley of Mexico until July 1, 2020, we selected only the 48,565 records with positive result and from Mexico City. We were only interested in the fields containing the name of the municipality and the number of positive coronavirus cases. We complemented our dataset with the population by municipality from Wikipedia.

We got the geographic coordinates of the Mexican states and municipalities in CDMX from *geocoder*. In any case, minimal corrections in the coordinates were required.

3 Methodology

The COVID-19 pandemic is caused by a completely novel coronavirus whose level of spread and affectations among different populations are still mostly unknown. Different countries around the

	Positive	Hospitalized	Dead
Diabetes	16.69	30.84	37.22
COPD	1.98	3.90	5.36
Asthma	2.96	2.59	2.33
Immunosup.	1.62	2.86	3.27
Hypertension	20.34	34.23	42.34
CVD	2.58	4.51	5.76
Obesity	19.92	23.83	25.45
CKD	2.39	5.06	7.04
Smoking	8.11	8.57	9.10

Figure 3: Percentage of comorbidities prevalence within a given condition

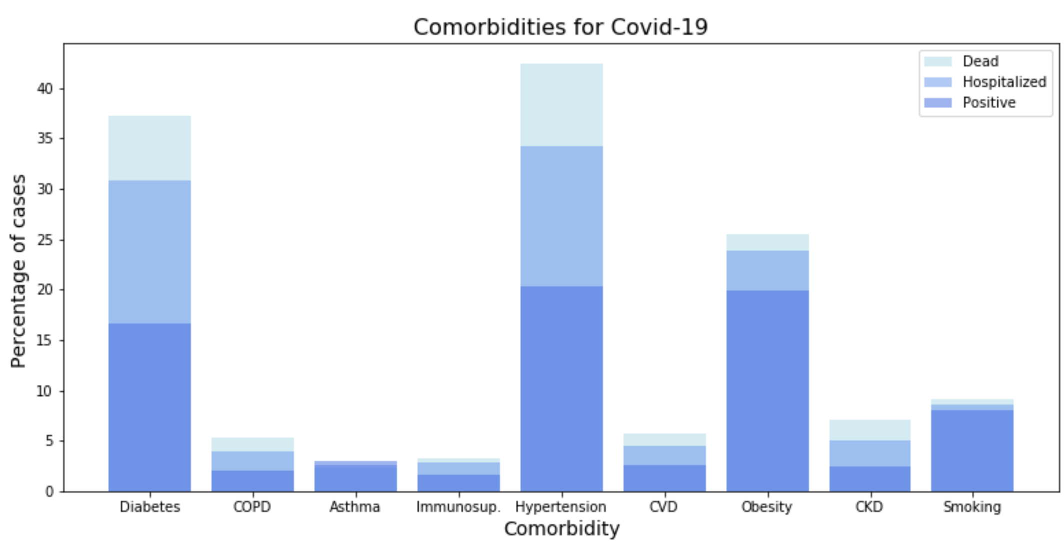

Figure 4: Comorbidities associated to COVID-19 pandemic in Mexico

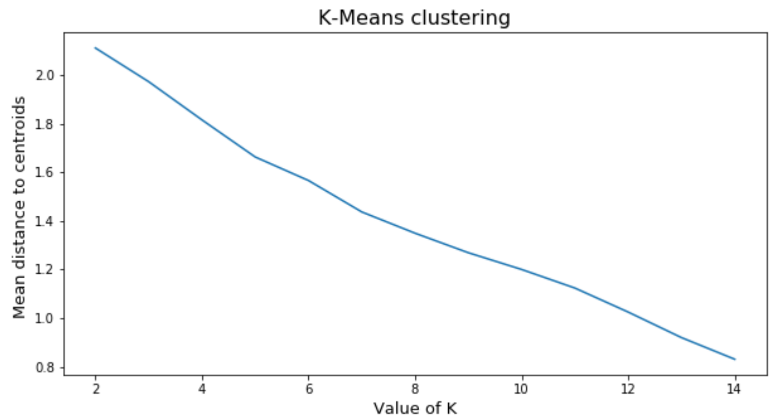

Figure 5: Determining the number of clusters

world took different actions to reduce the impact of the pandemic in their societies according to their particular conditions and limitations. In Mexico, the federal government decided to restrict the number of coronavirus tests since the beginning of the pandemic. This fact makes it difficult to have enough data to know the current state of the pandemic in the country and try to predict the evolution of it. Another particularity of the new coronavirus pandemic in Mexico is the prevalence of comorbidities associated to the coronavirus disease. Obesity, diabetes, hypertension and any other comorbidity related to them have a high incidence among Mexican population. For this reason, we observe particular phenomena in the Mexican case, such as health complications among younger individuals with COVID-19 in comparison with other countries (see Figure 2). We aim to use the K-means clustering method to group the 32 states in Mexico into clusters according to the prevalence of the different comorbidities associated to COVID-19.

3.1 Comorbidities of COVID-19 by state in Mexico

From the nine comorbidities realated to the new coronavirus reported in the original dataset we would decide which ones have more relevance to take into account for our clustering analysis. In table 3, we show the percentage of prevalence of each comorbidity within a given condition (positive test, hospitalized or dead). The three comorbidities with the highest prevalence among the Mexican population are hypertension, obesity and diabetes, which also represent an important risk factor when combined with the coronavirus disease. However, other comorbidities with a lower presence among general population in Mexico such as COPD, CVD and CKD represent a higher risk among population within such condition as can be seen in Table 3 and Figure 4. For instance, CKD has a prevalence of 2.39% among all population with a COVID-19 positive diagnosis, however this percentage is more than double within hospitalized patients and almost triple among the coronavirus patients who died.

According to the data we have, smoking is a comorbidity with a lower level of affectation among coronavirus patients in Mexico. On the other hand, asthma is the only comorbidity in the records for which the prevalence percentage decreases as the condition becomes more severe (the percentages are 2.96%, 2.59% and 2.33% for positive test, hospitalized and dead patients, respectively). This in an unexpected result since asthma could be a greater risk for many severe diseases but in the case of coronavirus there is not evidence supporting this as can be checked in the AAAAI report on *https://www.aaaai.org/conditions-and-treatments/library/asthma-library/covid-asthma*. For these

Alcaldia	Population	Hospitals	Positive_covid-19
Álvaro Obregón	727034	48	3395
Azcapotzalco	414711	34	2771
Benito Juárez	385439	48	1699
Coyoacán	620416	49	3065
Cuajimalpa de Morelos	186391	27	1045
Cuauhtémoc	531831	50	2892
Gustavo A. Madero	1185772	50	6167
Iztacalco	384326	28	2555
Iztapalapa	1815786	26	8332
Magdalena Contreras	239086	48	1287
Miguel Hidalgo	372889	48	1938
Milpa Alta	130582	2	1536
Tláhuac	360265	4	2430
Tlalpan	650567	49	3768
Venustiano Carranza	430978	49	2467
Xochimilco	415007	18	3218

Figure 6: Population and COVID-19 positive cases per municipality in CDMX

reasons, we decided to discard smoking and asthma from the comorbidities in our analysis.

As mentioned before, there has been a small number of tests for coronavirus in Mexico, hence we consider that it is more representative to center our analysis on the cases with lethal outcome (dead status in our dataset). Once we have determined which comorbidities to consider in the clustering process, we aim to find the optimal number of clusters in our analysis. First we normalize data using Z-score normalization, and then compute the mean distance to centroids for distinct values of k in K-means clustering. In Figure 5 we see that $k = 5$ is an adequate number of clusters according to the elbow method.

3.2 Hospital availability by municipality in CDMX

Our analysis on hospital availability in CDMX is based on a data frame containing population, number of coronavirus-positive tested persons and an estimate of hospitals and medical clinics of each municipality in Mexico City obtained from Foursquare (see Table 6). Using linear regression it is possible to model the relationship between the population and the number of coronavirus positive tests in each municipality, as expected given the nature of the viral transmission phenomenon (in Figure 7 we provide a graphical representation of this correlation). The Pearson correlation coefficient and p-value derived from data are 0.9781 and $5.9905 \times e^{-11}$, showing a strong correlation between population and coronavirus cases per municipality. On the contrary, population and number of hospitals per municipality have weak correlation values. This suggests that the distribution of health institutions in CDMX attends to socioeconomic reasons distinct from the total population of a given municipality.

We aim to use K-means clustering to group the 16 municipalities in Mexico City using the fields:

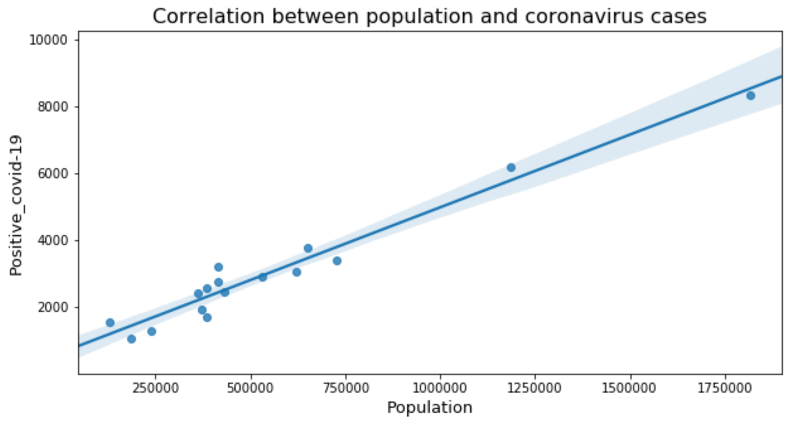

Figure 7: Correlation between population and coronavirus cases per municipality in CDMX

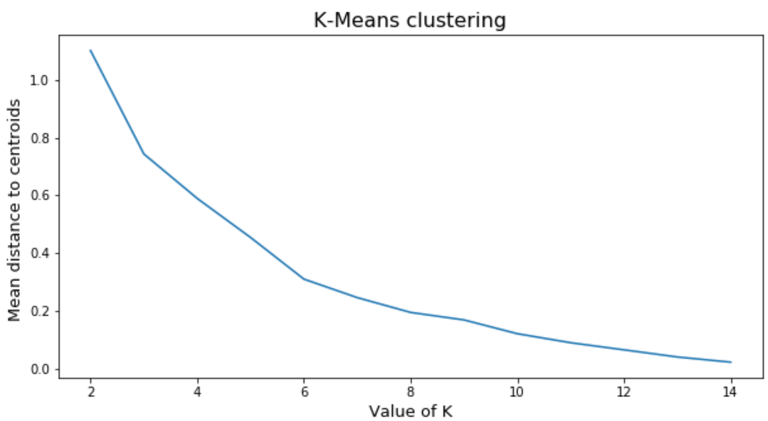

Figure 8: Determining the number of clusters

population, number of COVID-19 positive cases and number of hospitals and medical clinics. To find a adequate number of clusters we used the elbow method. After normalizing data, we computed the average distance to centroids for different values of number of clusters k (see Figure 8). From this analysis it follows that $k = 6$ is a good number of clusters according to the elbow method.

At this point, we want to emphasize that the number of hospitals per municipality we used here is a rough estimation. To find this number, we performed a Foursquare search using the category id corresponding to hospital venue in a radius of $2km$ around the administrative center of each municipality. There are some places under this search which do not correspond precisely to a hospital or medical clinic (for instance, it could be a clinical analysis laboratory). Nevertheless, we consider the result of this search a reasonable indicator of health care institutions in a given municipality.

Figure 9: Mexican states grouped by affectations of comorbidities

4 Results

4.1 Comorbidities of COVID-19 by state in Mexico

From the clustering of the Mexican states into groups according to the prevalence of comorbidities (Figure 9), we got the following results:

Cluster 1

States: Baja California, Campeche, Coahuila, Chiapas, Chihuahua, Querétaro, Quintana Roo, Sinaloa, Sonora, Tabasco, Tamaulipas, Veracruz and Yucatán.

Description: High prevalence of hypertension and relatively low prevalence of COPD.

Cluster 2

States: Ciudad de México, Colima, Estado de México, Guerrero, Hidalgo, Michoacán, Morelos, Oaxaca, Puebla and Tlaxcala.

Description: Low prevalence of hypertension and relatively low prevalence of CVD and CKD.

Cluster 3

States: Durango, Guanajuato, Jalisco, Nuevo León and San Luis Potosí.

Description: High prevalence of diabetes and hypertension and relatively high prevalence of COPD.

Cluster 4

States: Baja California Sur, Nayarit and Zacatecas.

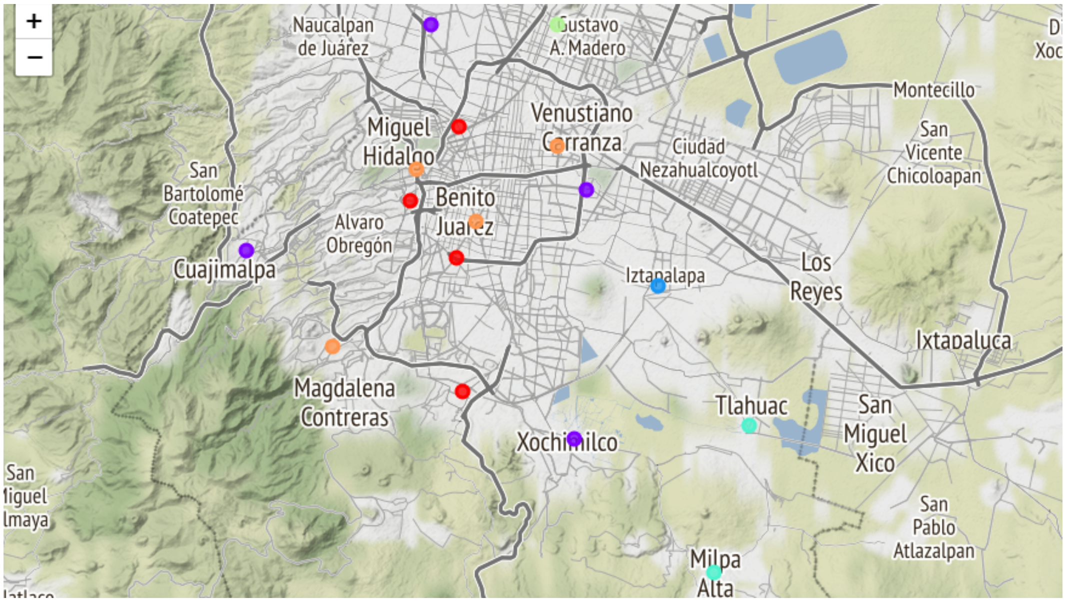

Figure 10: Hospital availability in CDMX by municipality

Description: High prevalence of immunosuppression and hypertension and relatively high prevalence of COPD, CVD and obesity.

Cluster 5

States: Aguascalientes.

Description: Extremely high prevalence of CKD and relatively high prevalence of immunosuppression and CVD. This cluster contains an outlier in data.

4.2 Hospital availability by municipality in CDMX

In the clustering process of the municipalities in Mexco City we considered the fields population, number of coronavirus positive cases and estimate of hospitals in each municipality. As we noticed in the previous section, there is a strong correlation between the total population and the coronavirus cases in each municipality, hence the groups we got are characterized by the ratio between the estimate of hospitals and the number of coronavirus positive cases or the total population per municipality. The groups ordered by this ratio are the following:

Cluster 1

Municipalities: Benito Juárez, Magdalena Contreras, Miguel Hidalgo and Venustiano Carranza.

Cluster 2

Municipalities: Álvaro Obregón, Coyoacán, Cuauhtémoc and Tlalpan.

Cluster 3

Municipalities: Azcapotzalco, Cuajimalpa, Iztacalco and Xochimilco.

Cluster 4

Municipalities: Gustavo A. Madero.

Cluster 5

Municipalities: Iztapalapa.

Cluster 6

Municipalities: Milpa Alta and Tláhuac.

5 Discussion

There are some relevant facts that can be deduced from the analysis of the dataset on COVID-19 pandemic in Mexico. First, smoking and asthma do not increase considerably the risk of health complications among coronavirus patients in Mexico as could be expected. Moreover, among asthma suffers the percentage of individuals that required hospitalization or died decreased in comparison with those with only a positive COVID-19 diagnosis. The main comorbidities for COVID-19 among Mexican population are hypertension, diabetes and obesity, however, comorbidities with lower prevalence in the country such as CKD, CVD or COPD represent a higher risk for coronavirus patients under these conditions.

The clustering analysis of the Mexican states according to the comorbitidity affectations revealed that the clusters keep some relation with geographical proximity. This fact is particular observable in cluster 2, which only contains states in the center of the country, or the cluster 2, which mainly contains sates in the north and southeast. We have not enough data to establish the reason for this relation.

From the analysis of hospital availability in Mexico City it is possible to establish a strong correlation between the total population and the coronavirus pandemic spread in each municipality. On the other hand, we noticed an unbalanced distribution of hospitals with relation to population in the different municipalities in the city. From the clustering analysis, we got results that agreed with what was expected. There are more health institutions in the central part of the city and this availability decreases as we move away form the center. We also detected three municipalities where the hospital availability is highly deficient in comparison with the reported number of coronavirus cases: Milpa Alta, Tláhuac and Iztapalapa.

6 Conclusion

The COVID-19 pandemic in Mexico is having a particular evolution according to the particular conditions in the country. In this work, we wanted to analyze two important problems related to the pandemic in Mexico: the prevalence of main comorbidities in different states and the hospital availability in Mexico City. In both cases we used the K-means clustering method to identify similarities among the Mexican states and the municipalities in Mexico City with respect to comorbidities prevalence and hospital availability, respectively. We detected hypertension, diabetes, obesity, CKD, CVD, COPD and immunosuppression as the main comorbidities associated to COVID-19 in Mexico and how their prevalence varies from one state to other. We also analyze the hospital distribution in the municipalities in CDMX in comparison with the coronavirus cases per municipality. The results we obtained may be useful for governments (local or federal) to improve their actions against the coronavirus pandemic. These results may also be of interest to individuals suffering from one or more comorbidities.