Number of hospitals of the biggest cities around the world

Thiago Bodruk

March 25, 2020

1. INTRODUCTION

1.1 Background and motivation

Given the recent COVID-19 global pandemic, several major cities around the world will face infrastructure collapse related to insufficient healthcare, hospitals, and emergency services. At present, the Coronavirus has made thousands of victims around the globe. The World Health Organization reported the following statistics related to the pandemic:

- On average, 15% of the infected people need medical health care, with symptoms like fever, cough, chest discomfort, aches, and stuffy nose;
- On average, 5% of the infected people need intensive health care, including segregation and artificial breathing, presenting symptoms like shortness of breath, fever, coughs and chest discomfort;
- On average, between 2% and 5% of the infected people may occasionally die due to respiratory complications, infections, and pneumonia.
- People older than 65 years old, smokers, patients with respiratory problems, diabetes or immunosuppressed belong to the risky group and, have higher odds of serious complications and death.

1.2 Problem

Based on those statistics, a city with 10 million citizens probably will have 1,500,000 infected people, who will need specialized healthcare and, in specific cases, intensive health care. Furthermore, between 200,000 and 500,000 infected people may die due to serious breathing complications.

Are those cities prepared to handle pandemics and catastrophic events? How many hospitals and intensive care services are in major cities around the world? What is the average citizen per hospital rate? How do those cities compare to each other? How are the number of hospitals and the population size related?

1.2 Interest

Naturally, the medical and pharmaceutical industries invest huge amounts of money every year to provide high-quality products and services to the population. Business intelligence teams or data-driven companies will benefit a lot with the list of potential zones, clusters, and other comparative analysis.

2. DATASETS

2.1 Data sources

The geographic and demographic data of the biggest cities in the world were extracted from Wikipedia's related page. To complete the location features, the latitude and longitude coordinates were obtained using Python's Geopy library. Finally, to perform the map plots and the comparative analysis, the list of hospitals of each city was extracted using Foursquare's API.

2.2 Data cleaning

The data scraped from Wikipedia was extracted and converted to a dataframe using Pandas. This data contained invalid characters and annotations throughout the source table. First, I converted the original multi-index headers to a single-index version. Then I removed every non-numeric character and converted the numeric columns to integer numbers. Finally, I double-checked all the coordinates given by Geopy's library plotting them into Folium maps to verify the precision of the provided. In Moscow, the location provided by Wikipedia was slightly wrong, plotting the city center far away from the actual center. To fix that, I inserted the correct locations manually, using coordinates provided by Google Maps.

2.3 Feature selection

Combining the data extracted from Wikipedia, Geopy and Foursquare's API, after the data cleaning, there were 78 cities and 14 features on the data. The features are listed below:

- City name;
- Country;
- City definition;
- City population;

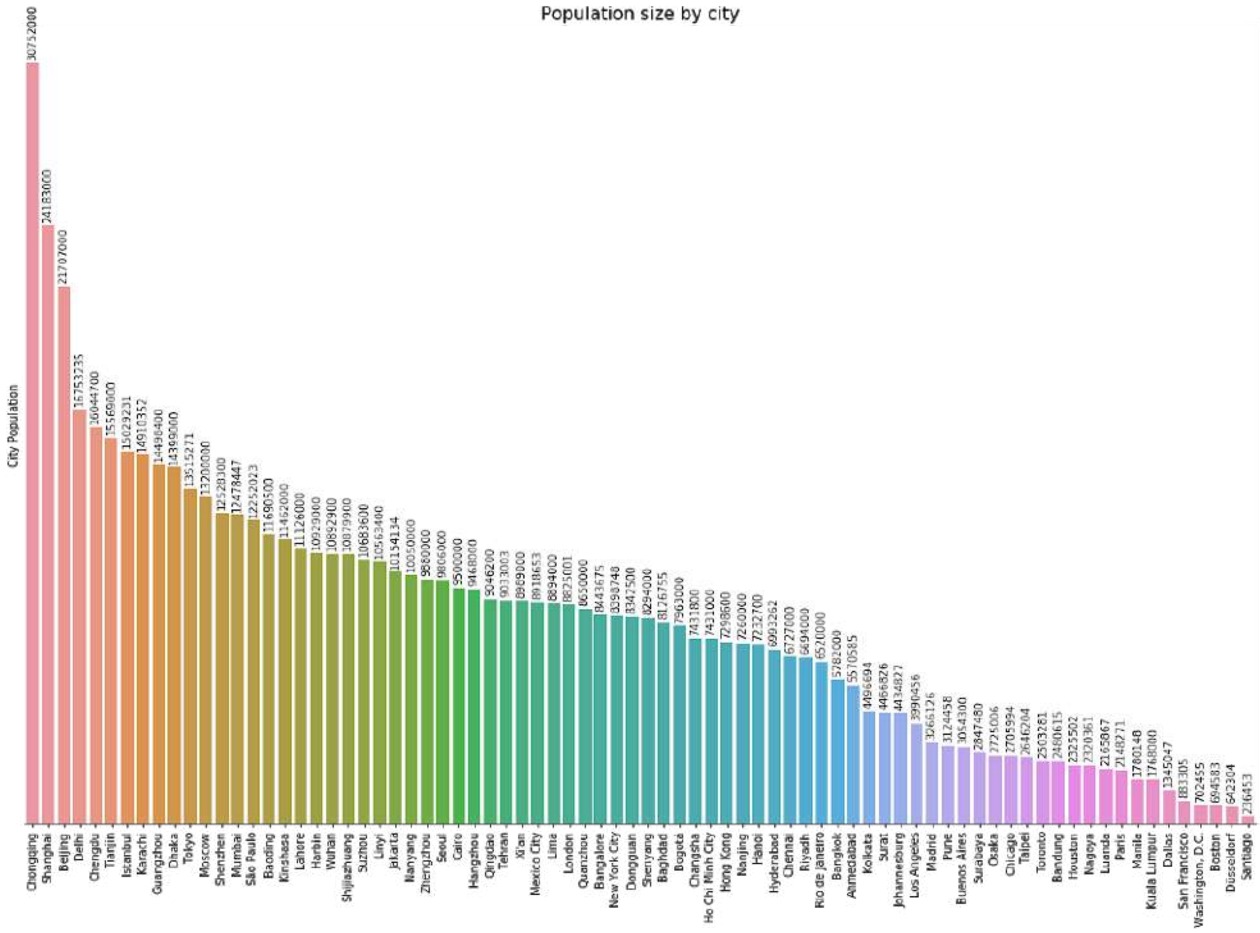
- City area (Km^2);
- Metropolitan population;
- Metropolitan area (Km^2);
- Urban population;
- Urban area (Km^2);
- Demographic density (inhabitants per Km);
- Latitude;
- Longitude;
- Number of hospitals;
- Inhabitants per hospital.

The *demographic density* and the *inhabitants per hospital* columns were created based on the other columns to provide a better understanding of the data. Some columns, like the metropolitan population and the metropolitan area, had missing information, so they were not used to evaluate the cluster analysis. All the other columns were used on different comparative analysis.

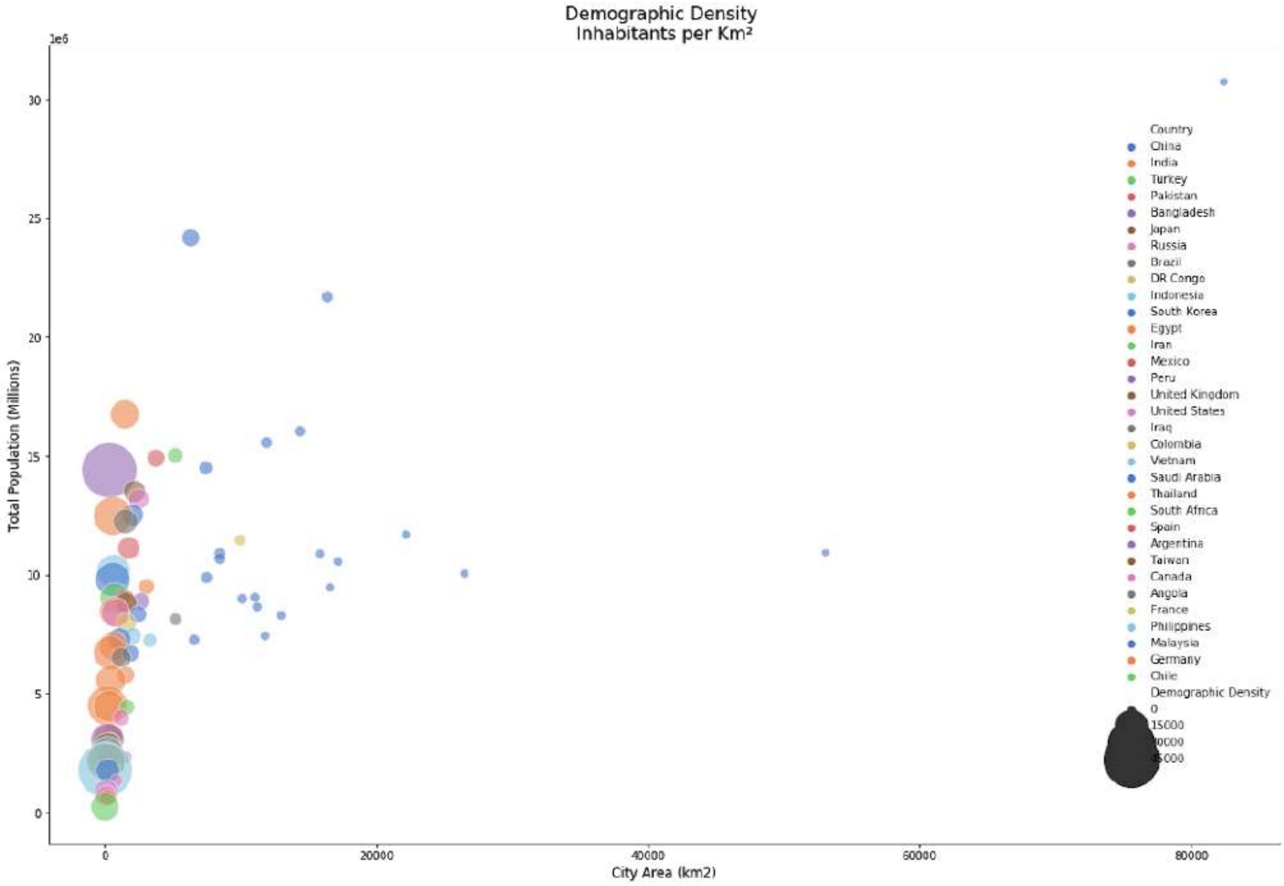
3. METHODOLOGY

3.1 Analyzing the biggest cities

The process started with two basic plots. First, a column chart to compare the cities according to the population size. The chart was sorted in descending mode, with the bigger cities on the left of the chart and the smaller cities on the right side.



Second, a scatter plot to help understand the demographic density, plotting the relationship between the city areas and the population size. Each city was plotted on a different color, grouped by the country. The size of the circles represents the demographic density of each city.



3.2 Selecting the biggest cities

For a better understanding of the cities dataframe, a new dataframe was created, containing the number of cities per country. This list was generated to figure out the countries that have the biggest cities on the planet.

	City Population	City Area(km2)	Demographic Density	Cities
Country				
China	295631800	384199	37675.494979	24
India	69054182	5144	129363.584075	9
United States	21046090	5448	38354.642105	8
Japan	18560638	2742	25397.349493	3
Indonesia	15482229	1183	38170.414591	3
Pakistan	26036352	5552	10223.318605	2
Brazil	18772023	2742	13395.127286	2
Vietnam	14663700	5386	5780.779416	2
Canada	2503281	630	3973.461905	1
South Africa	4434827	1645	2695.943465	1

Table - Number of cities per country

After the countries dataframe was created, a new column chart was plotted to help visualize the top 10 population size by country, considering the grouped cities list.

Finally, I selected the biggest city from each country of the previous list. So, on the selected list, the following cities were selected:

City	Country	City Definition	City Population	City Area(km2)	Metropolitan Population	Metropolitan Area(km2)	Urban Population	Urban Area(km2)	Demographic Density
Zhengzhou	China	Special administrative region	30752000	82403	30752000	82403	22125000	4144	6611.050725
Surat	India	National capital territory	16753235	1484	29000000	7256	28125000	2240	21935.092683
Istanbul	Turkey	Metropolitan municipality	15029231	5196	15029231	5196	13860000	1360	2892.461701
Lahore	Pakistan	Metropolitan city	14910352	3780	14910352	3780	16900000	1036	6278.781038
Dhaka	Bangladesh	Capital city	14399000	338	14543124	338	18595000	453	42600.591716
Tokyo	Japan	Metropolis prefecture	13515271	2191	37274000	13452	38505000	8223	12111.137778
Moscow	Russia	Federal city	13200000	2511	13200000	2511	16555000	5698	5256.869773
São Paulo	Brazil	Municipality	12252023	1521	21734682	7947	20935000	3043	8055.241946
Kinshasa	DR Congo	City-province	11462000	9965	11462000	9965	12960000	583	1150.225790
Surabaya	Indonesia	Special capital region	10154134	664	33430285	7063	34365000	3367	15292.370482

Table - Selected cities

3.3 Analyzing the number of hospitals

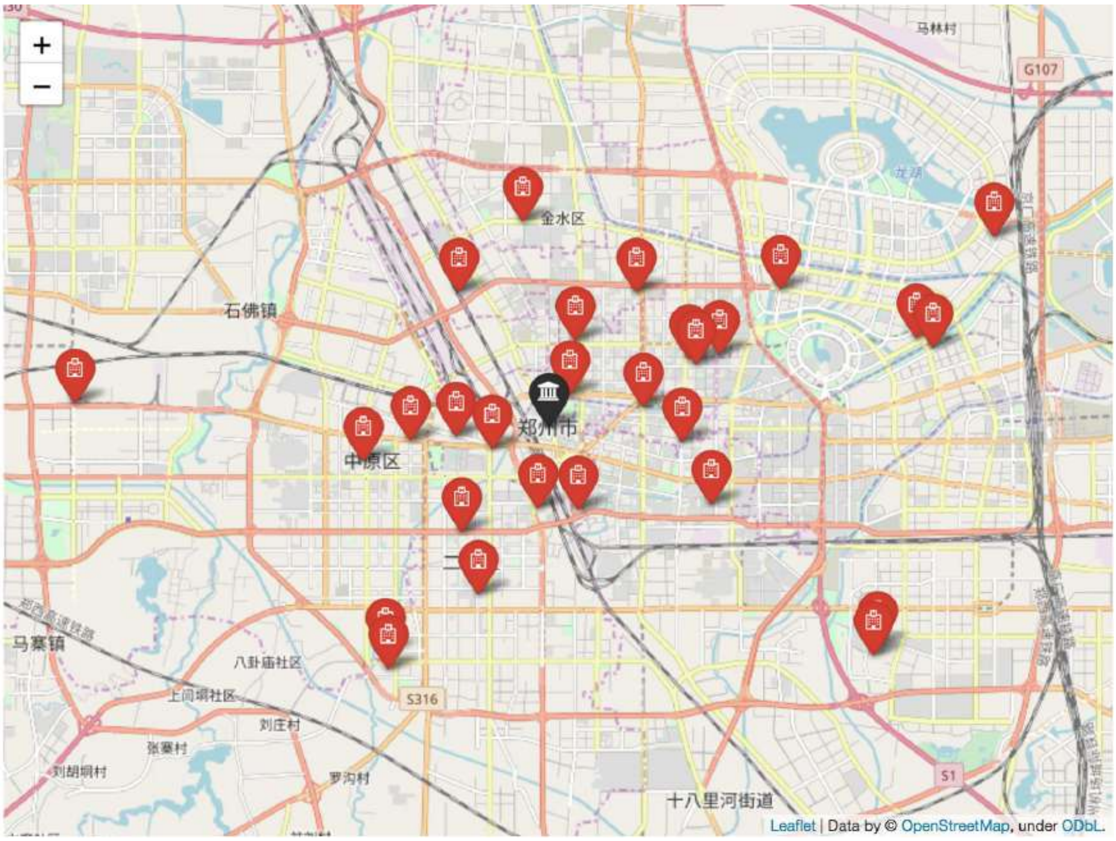
In the next step, an algorithm was created to get a list of hospitals in each city. Due to Foursquare API limitations, each list was limited to 50 venues per search in a radius of 50Km.

Afterward, a Folium map was plotted for each one of the ten cities, showing the location of their respective hospitals and the geographical center of each city.

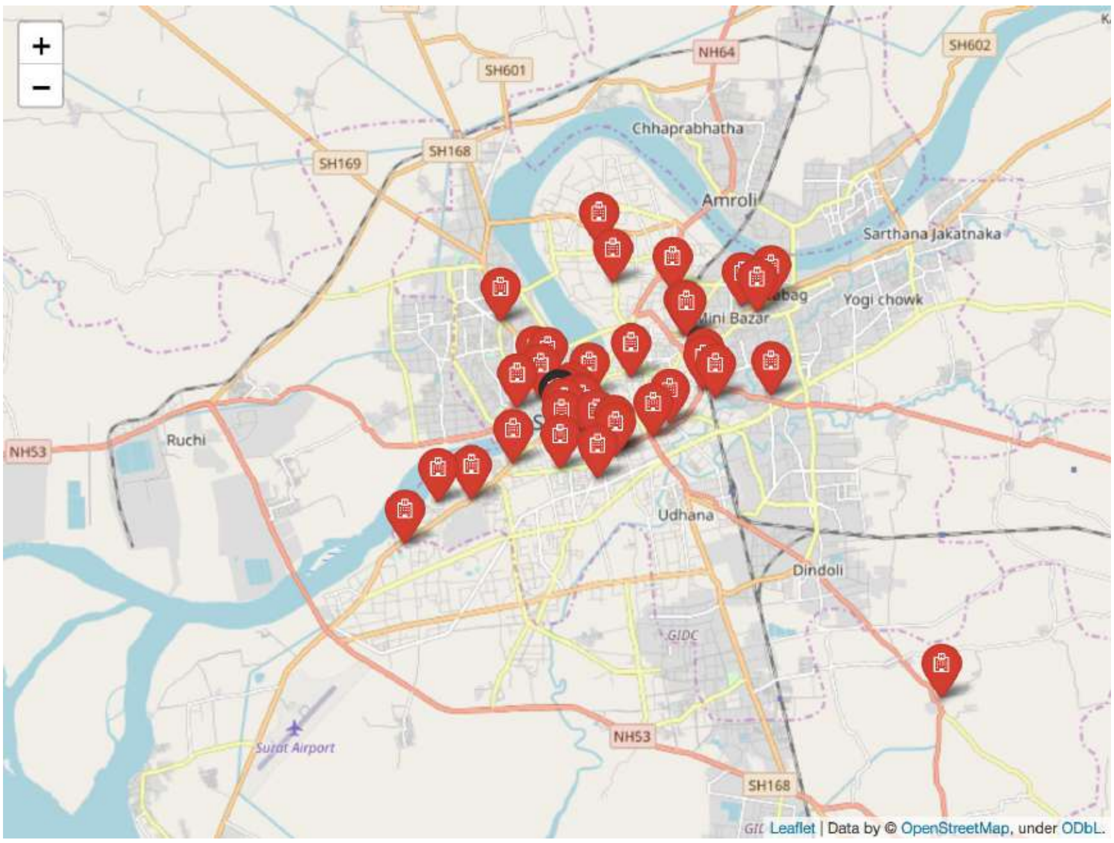
City	Country	City Definition	City Population	City Area(km2)	Metropolitan Population	Metropolitan Area(km2)	Urban Population	Urban Area(km2)	Demographic Density	Latitude	Longitude	Hospitals
Zhengzhou	China	Special administrative region	30752000	82403	30752000	82403	22125000	4144	6611.050725	34.759188	113.652408	30.0
Surat	India	National capital territory	16753235	1484	29000000	7256	28125000	2240	21935.092683	21.186461	72.808128	48.0
Istanbul	Turkey	Metropolitan municipality	15029231	5196	15029231	5196	13860000	1360	2892.461701	41.076602	29.052495	50.0
Lahore	Pakistan	Metropolitan city	14910352	3780	14910352	3780	16900000	1036	6278.781038	31.565608	74.314177	50.0
Dhaka	Bangladesh	Capital city	14399000	338	14543124	338	18595000	453	42600.591716	23.759357	90.378814	50.0
Tokyo	Japan	Metropolis prefecture	13515271	2191	37274000	13452	38505000	8223	12111.137778	35.682839	139.759455	50.0
Moscow	Russia	Federal city	13200000	2511	13200000	2511	16555000	5698	5256.869773	55.751244	37.618423	50.0
São Paulo	Brazil	Municipality	12252023	1521	21734682	7947	20935000	3043	8055.241946	-23.550651	-46.633382	50.0
Kinshasa	DR Congo	City-province	11462000	9965	11462000	9965	12960000	583	1150.225790	-4.321706	15.312597	22.0
Surabaya	Indonesia	Special capital region	10154134	664	33430285	7063	34365000	3367	15292.370482	-7.245972	112.737827	50.0

Table - Number of hospitals per city

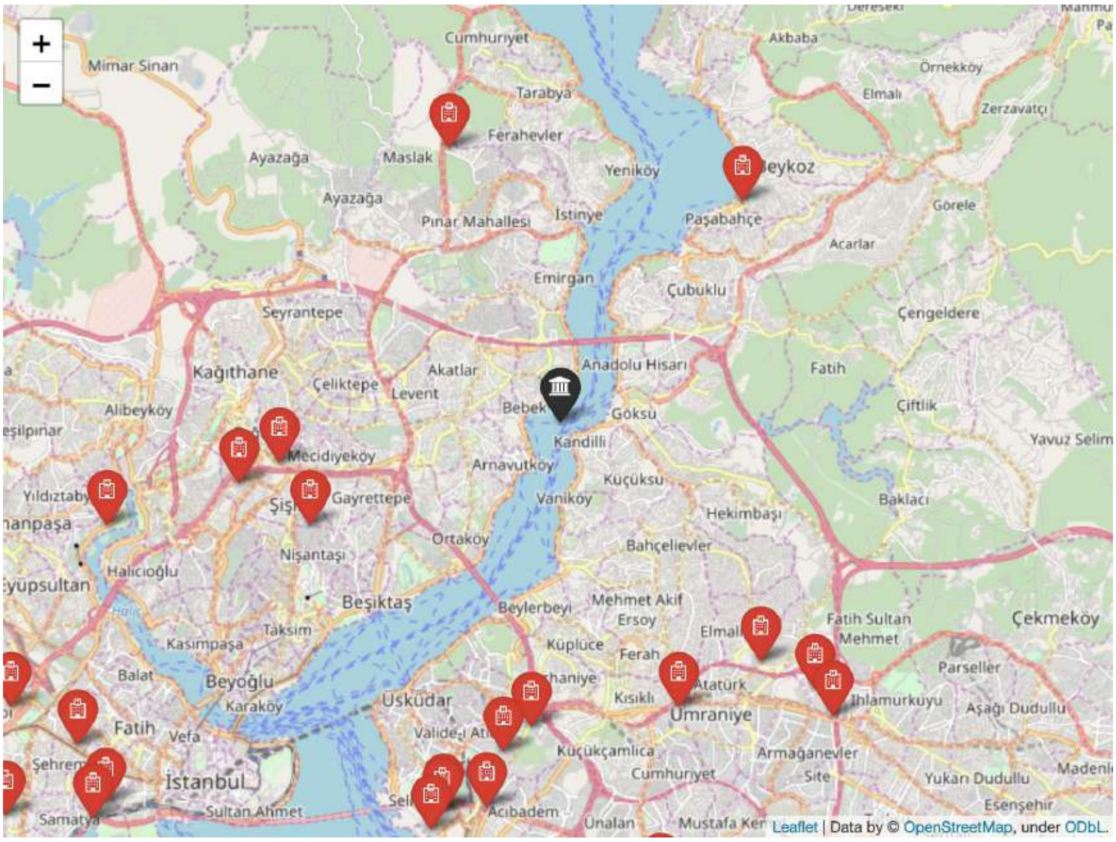
Zhengzhou, China



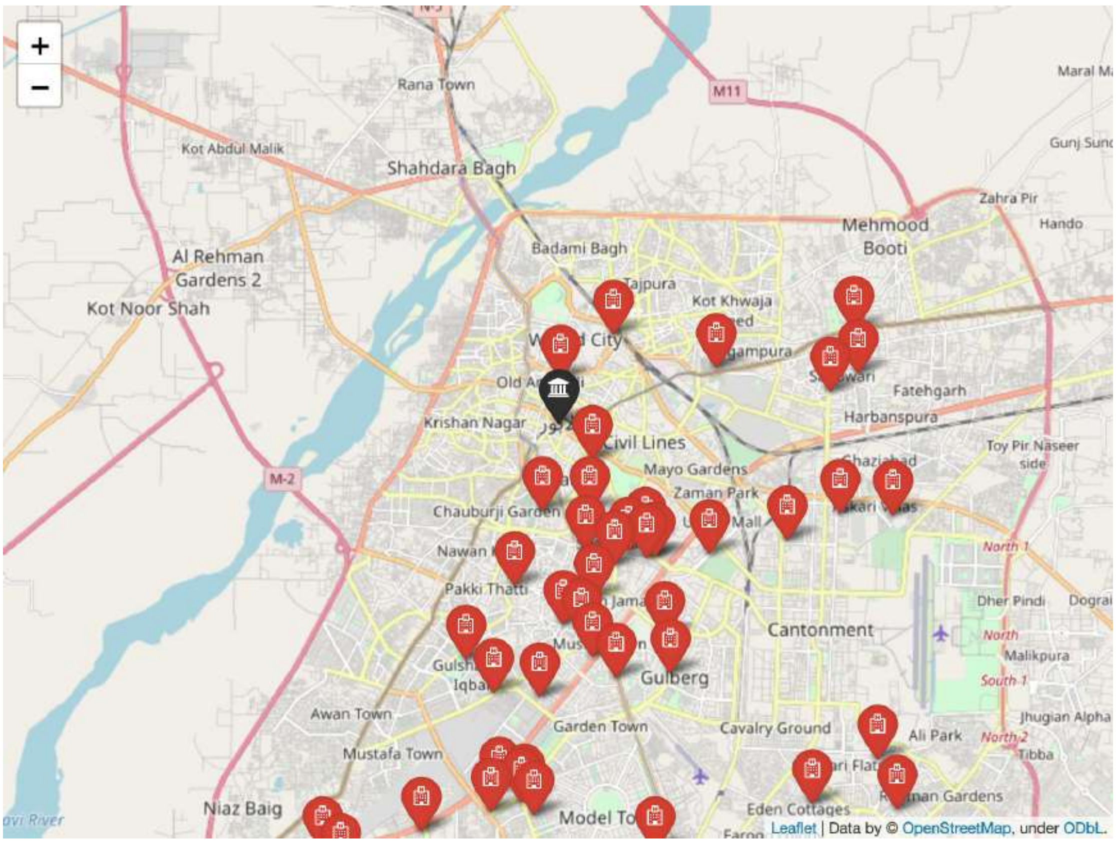
Surat, India



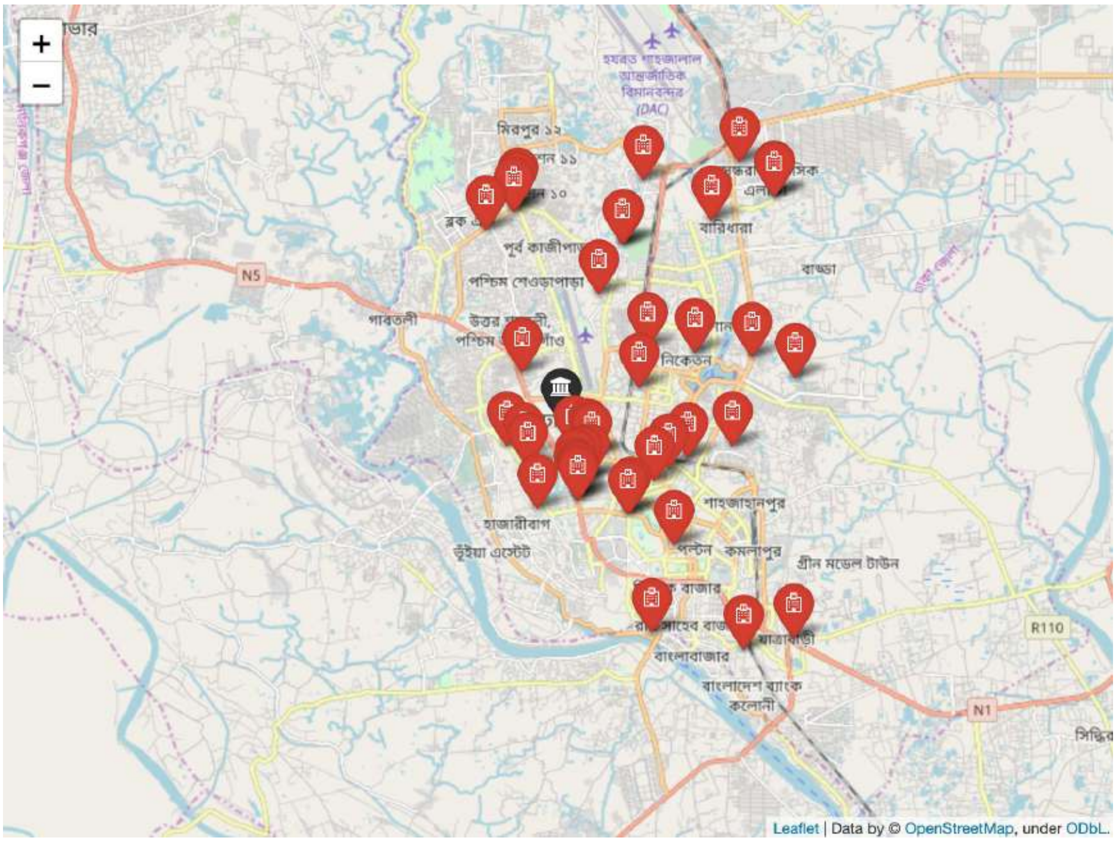
Istanbul, Turkey



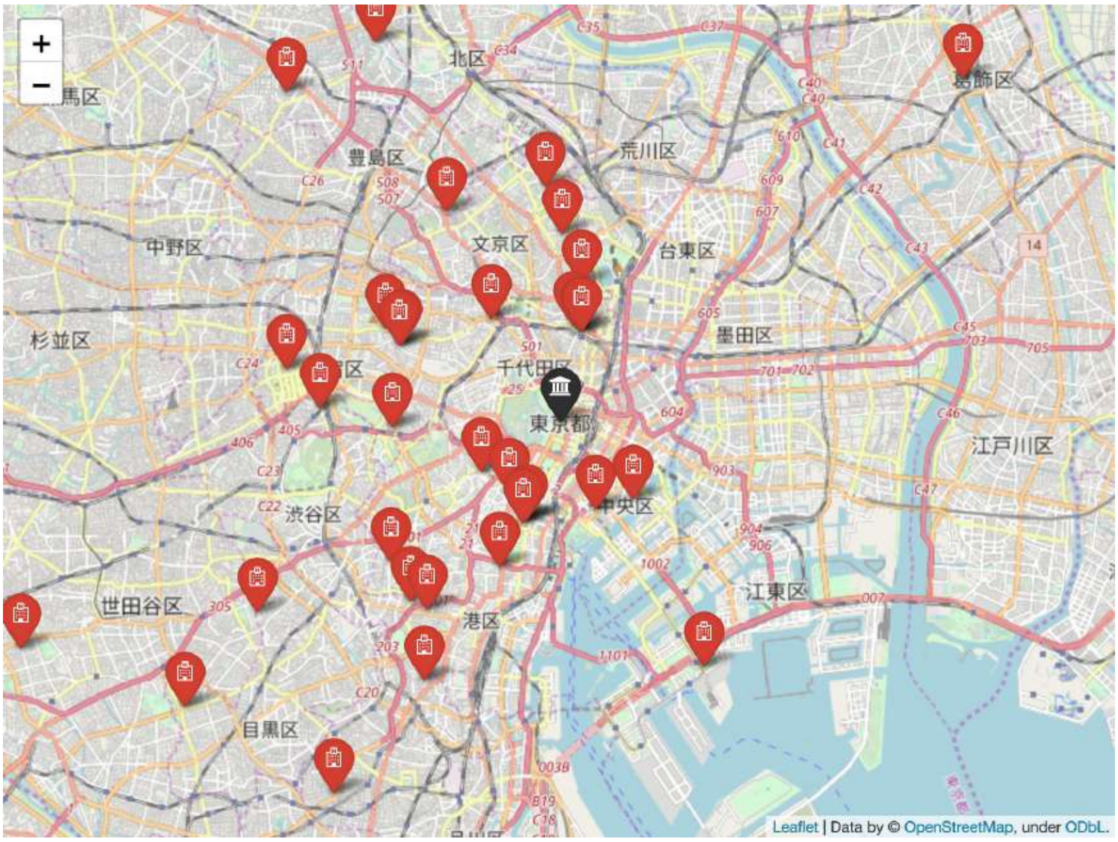
Lahore, Pakistan



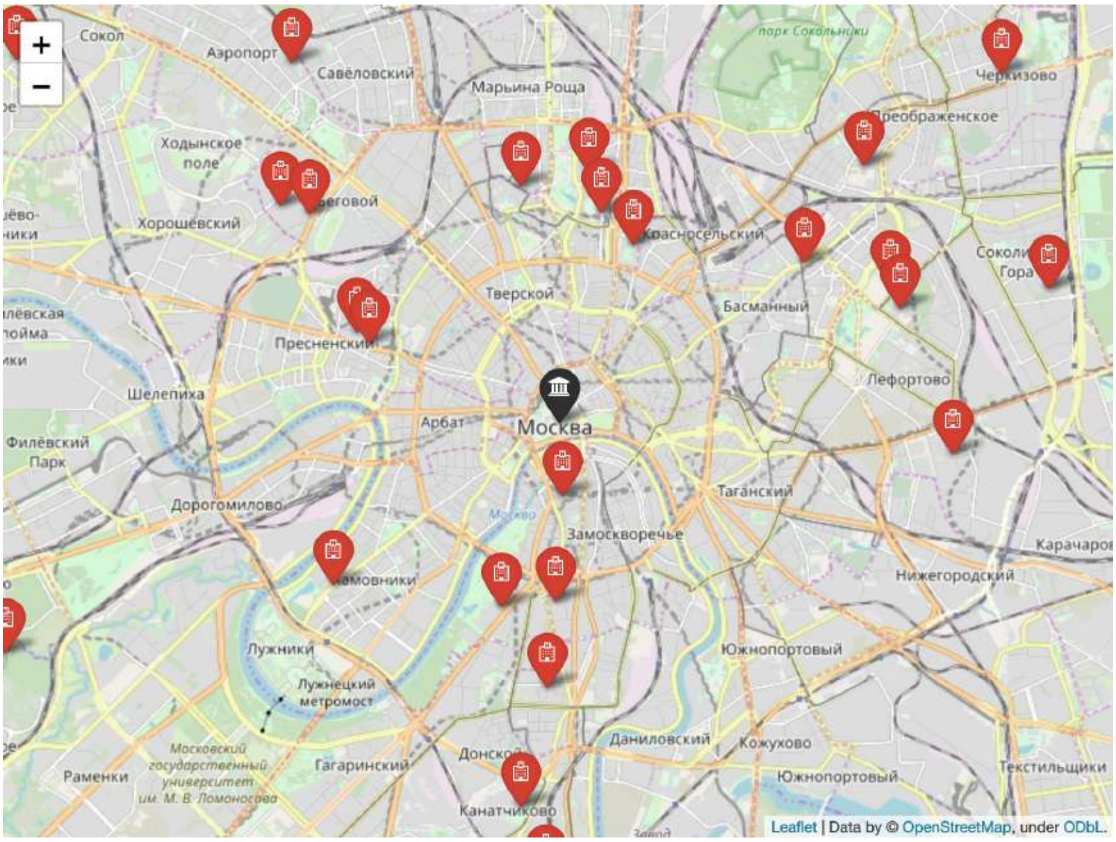
Dhaka, Bangladesh



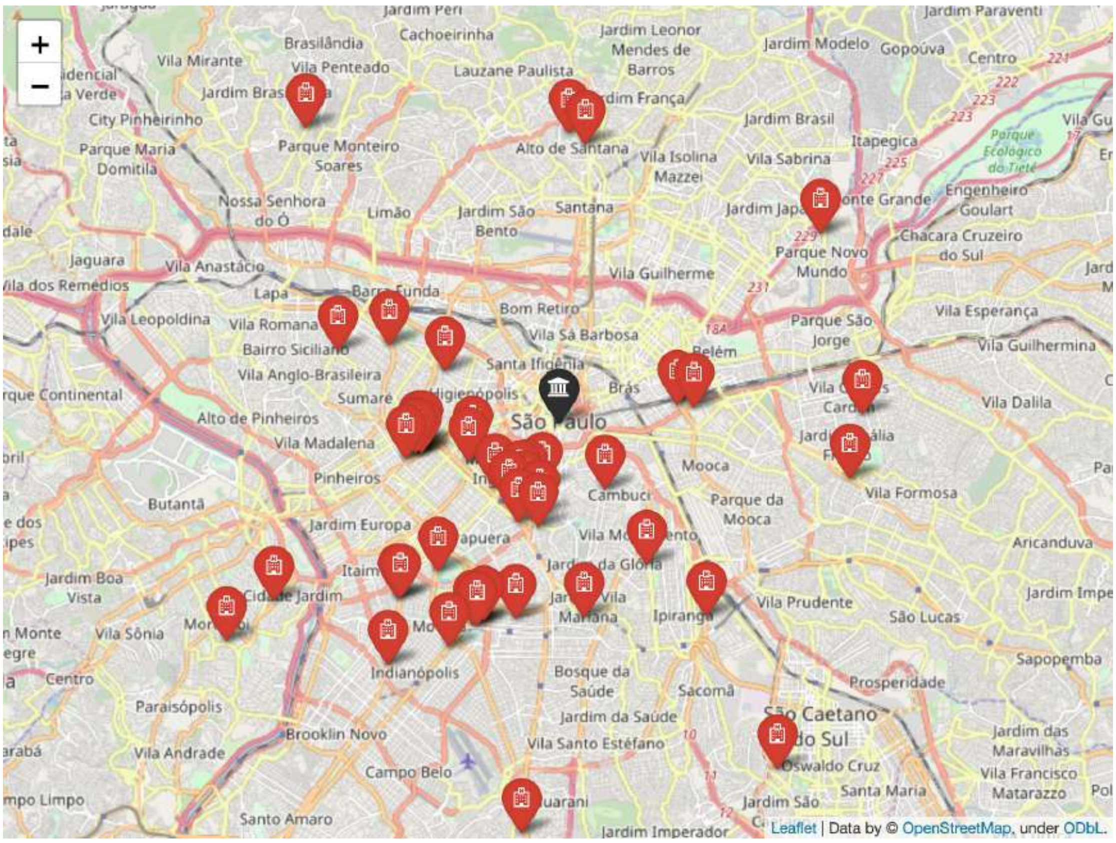
Tokyo, Japan



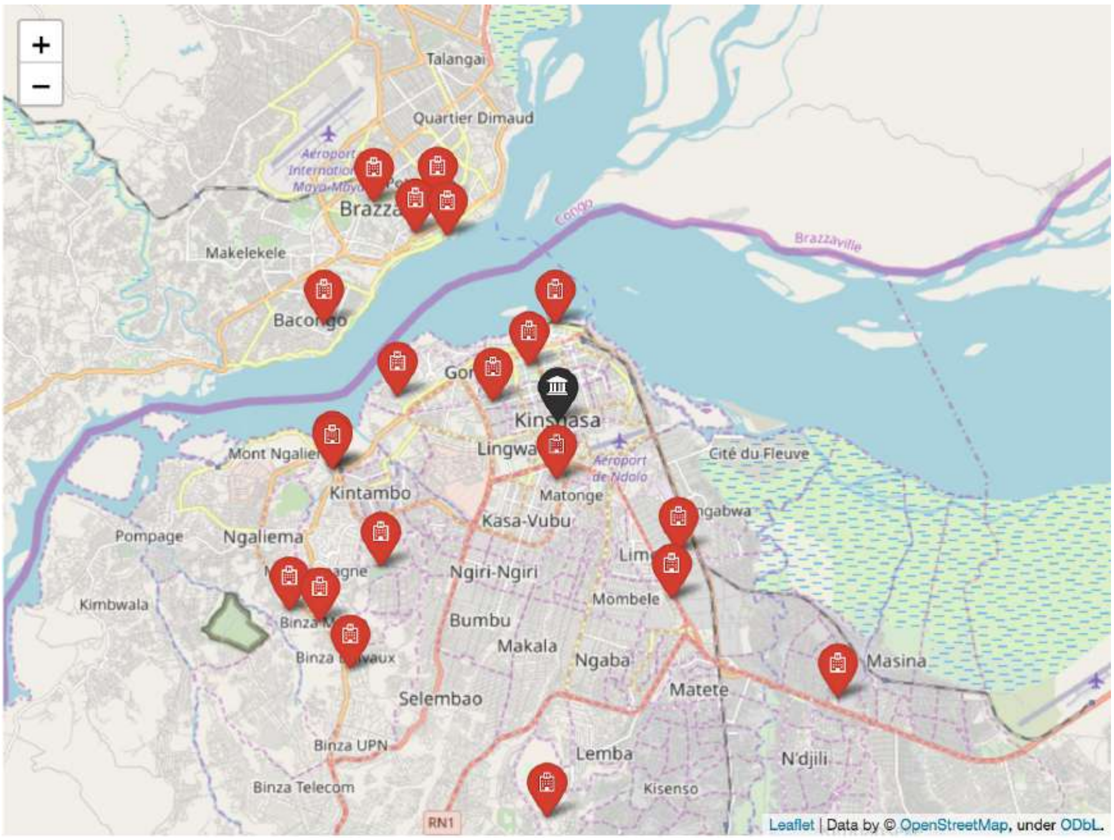
Moscow, Russia



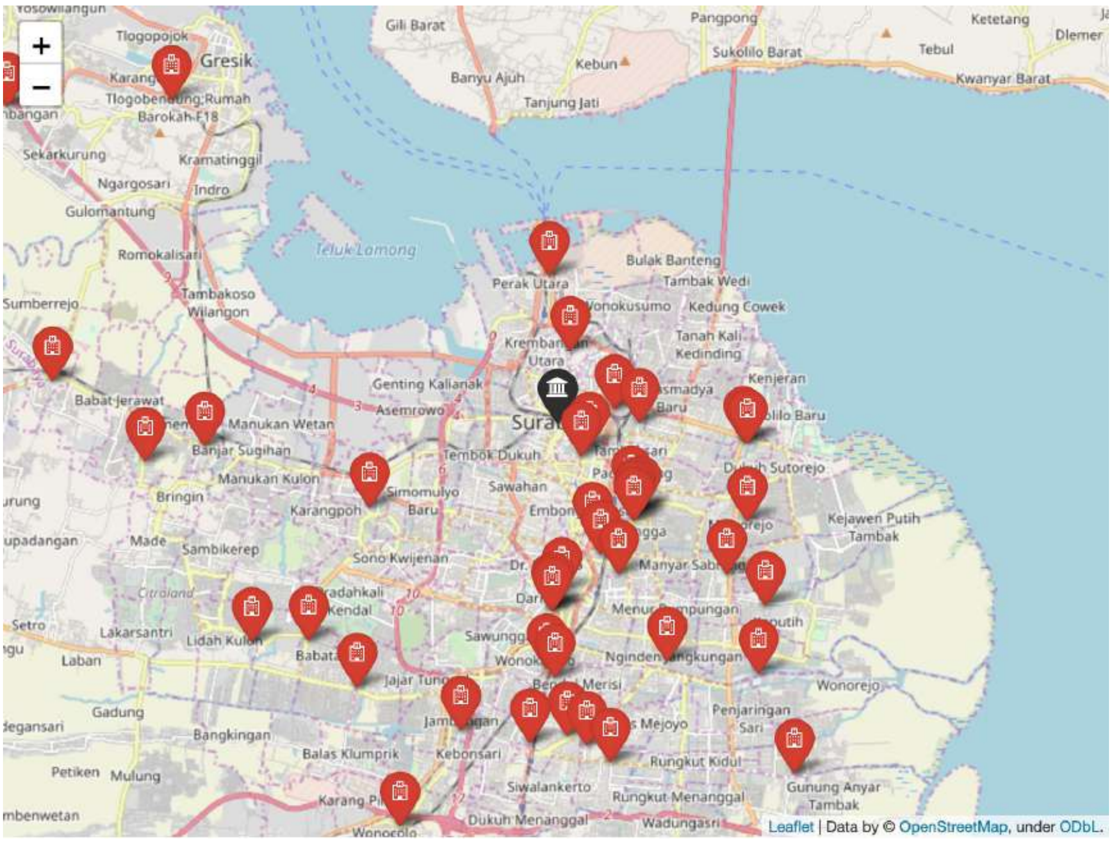
São Paulo, Brazil



Kinshasa, DR Congo

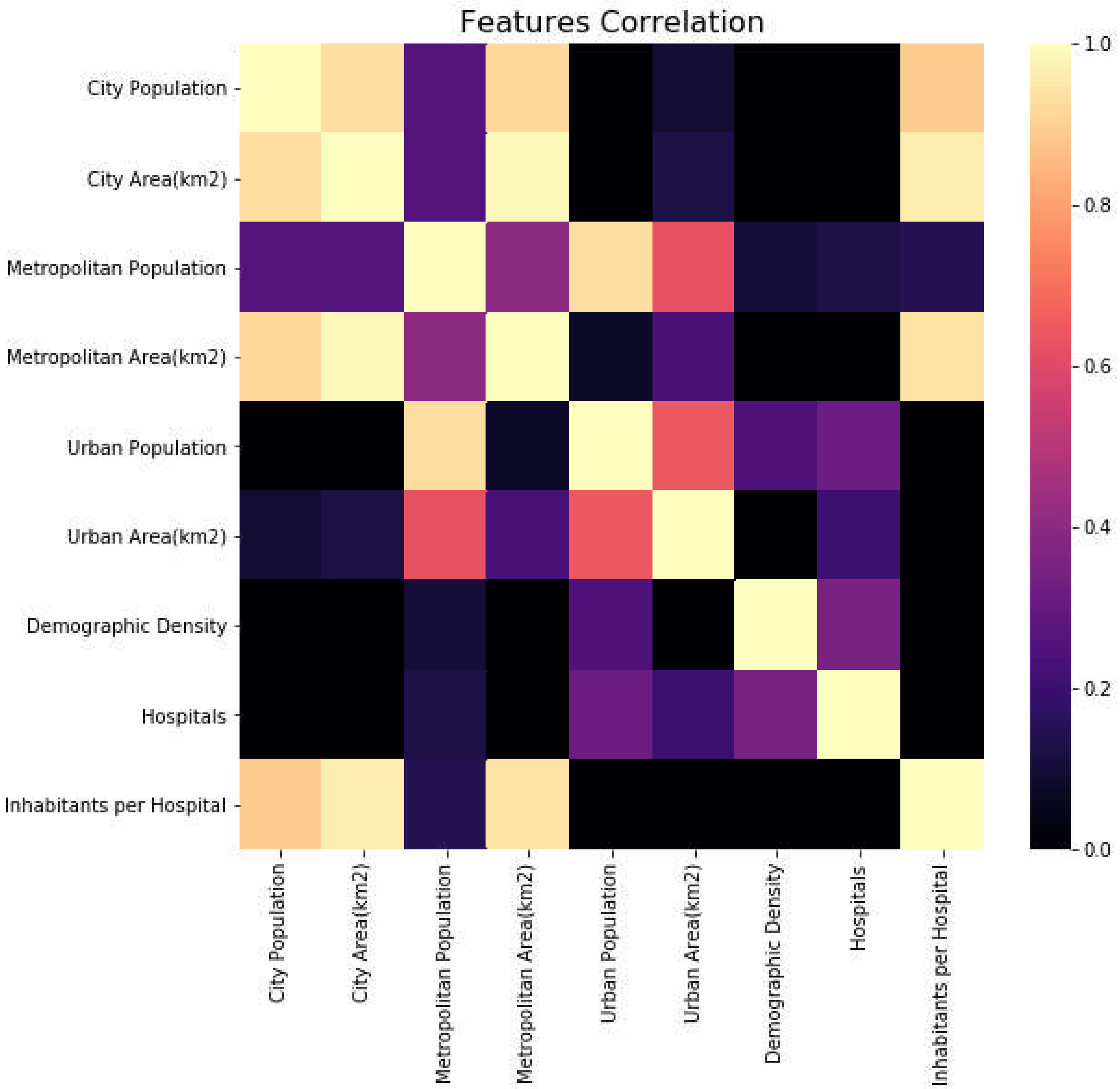


Surabaya, Indonesia



3.4 Features Correlation

Before the clustering process, a correlation model was created to evaluate the relationship between the different features, focusing on the optimal features. Then, a heatmap was plotted to help visualize the strength of each relationship.



As seen, we can notice a strong relationship between population and area size. Furthermore, we can see a discrete relationship between population size and the number of hospitals.

3.5 Cluster Model

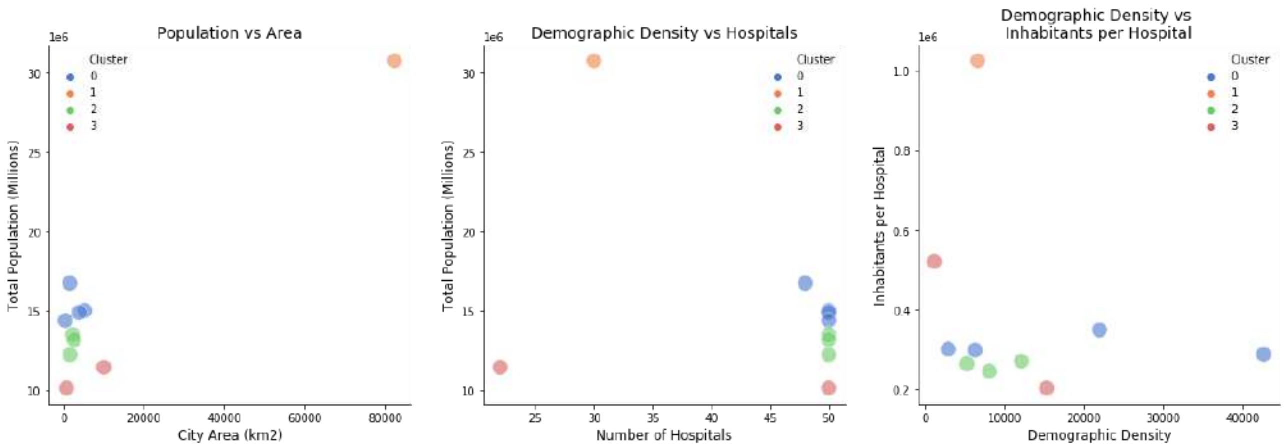
Finally, a cluster model was created using the K-means method, considering a number of clusters equals to four. To build the cluster model, the following features were selected: city population, city area (km^2), demographic density, number of hospitals, inhabitants per hospital. Due to miss information on the metropolitan and urban columns, those columns were not included in the clustering model. Although, they are not so relevant, as seen on the correlation heatmap.

	Country	City Population	City Area(km2)	Demographic Density	Hospitals	Inhabitants per Hospital	Cluster
City							
Zhengzhou	China	30752000	82403	6611.050725	30.0	1.025067e+06	1
Surat	India	16753235	1484	21935.092683	48.0	3.490257e+05	0
Istanbul	Turkey	15029231	5196	2892.461701	50.0	3.005846e+05	0
Lahore	Pakistan	14910352	3780	6278.781038	50.0	2.982070e+05	0
Dhaka	Bangladesh	14399000	338	42600.591716	50.0	2.879800e+05	0
Tokyo	Japan	13515271	2191	12111.137778	50.0	2.703054e+05	2
Moscow	Russia	13200000	2511	5256.869773	50.0	2.640000e+05	2
São Paulo	Brazil	12252023	1521	8055.241946	50.0	2.450405e+05	2
Kinshasa	DR Congo	11462000	9965	1150.225790	22.0	5.210000e+05	3
Surabaya	Indonesia	10154134	664	15292.370482	50.0	2.030827e+05	3

Table - Clustered cities

3.6 Clusters Comparison

Afterward, new scatter plots was created, comparing the cities, considering the processed clusters, among different correlations: population versus area; demographic density versus the number of hospitals; and demographic density versus inhabitants per hospital.



4. RESULTS

Analyzing the cluster model, we can identify evident patterns among the different cities. As mentioned in the methodology section, the population size and the area size had strong relevance on the clustering result. The four clusters are described as following below.

4.1 Bronze cluster

This cluster number #3 contains two cities: Kinshasa in DR Congo and Surabaya in Indonesia. Those cities have respectively 1 million and 1.1 million inhabitants and, a rate of 521.000 and 203.082 inhabitants per hospital. Compared to the other clusters, those cities were labeled as bronze, considering mainly the population size below 1.2 million.

4.2 Silver cluster

The cluster number #2 contains three cities: São Paulo in Brazil, Moscow in Russia and Tokyo in Japan. All the cities look very similar considering both population size, with 12.2 million, 13.2 million and 13.5 million inhabitants respectively; and area size range, from 1,521Km^2 to 2.511Km^2. In a similar way, all the cities have a similar rate considering the narrow range between 245,040 to 270,305 inhabitants per hospital. Compared to the other clusters, those cities labeled as silver, considering the strong similarity on population size, area size and inhabitants per hospital rate.

4.3 Gold cluster

This cluster number #0 contains four of the biggest cities in the world: Dhanka in Bangladesh, Lahore in Pakistan, Istanbul in Turkey and Surat in India. All of those cities have a population size between 14.3 and 16.7 million inhabitants. The area size has a wider range, starting from 338Km^2 to 5,196Km^2. Considering the inhabitants per hospital rate, the four cities have a range between 287,980 and 349,025 inhabitants per hospital. Compared to the silver cluster, the gold cluster has similar cities with a population size slightly bigger.

4.4 Platinum cluster

The cluster number #1 contains just one city: Zhengzhou in China, the biggest city on the planet. This city alone has an amazing population size of 30.7 million inhabitants; a city area of 82,403Km^2; and an inhabitant per hospital rate of 1,025,067. Despite the population size, the demographic density of 6,611 inhabitants per kilometer isn't the biggest of the list, when compared with cities like Dhaka and Surat, with 42,600 and 21,935 inhabitants per Km, respectively. Given its huge proportions, Zhengzhou is labeled as platinum, with double the population size of the second place on the list; and eight times the area size of the second biggest area of the list. Zhengzhou is, with no doubt, the biggest city on the planet.

5. DISCUSSION

The COVID-19, as well as other respiratory diseases, spreads very fast in big cities and high-density areas. The demographic density seems to have a strong relationship with the growth rate of the pandemic. In small cities, the odds of infection are low due to the lower risk of physical contact with multiple individuals during a daily routine.

Based on this principle, the higher the demographic density, the higher the risk of infection among the citizens. To handle this situation, the lesser the inhabitants per hospital, the higher the chance of emergency care, especially on pandemic cases.

	Country	City Population	City Area(km2)	Demographic Density	Hospitals	Inhabitants per Hospital	Cluster
City							
Surabaya	Indonesia	10154134	664	15292.370482	50.0	2.030827e+05	3
São Paulo	Brazil	12252023	1521	8055.241946	50.0	2.450405e+05	1
Moscow	Russia	13200000	2511	5256.869773	50.0	2.640000e+05	1
Tokyo	Japan	13515271	2191	12111.137778	50.0	2.703054e+05	1
Dhaka	Bangladesh	14399000	338	42600.591716	49.0	2.938571e+05	2
Lahore	Pakistan	14910352	3780	6278.781038	50.0	2.982070e+05	2
Istanbul	Turkey	15029231	5196	2892.461701	50.0	3.005846e+05	2
Surat	India	16753235	1484	21935.092683	41.0	4.086155e+05	2
Kinshasa	DR Congo	11462000	9965	1150.225790	22.0	5.210000e+05	3
Zhengzhou	China	30752000	82403	6611.050725	30.0	1.025067e+06	0

Table - Cities sorted ascending by inhabitants per hospital

5.1 Model Precision

According to the clusters model, Surabaya in Indonesia has the best hospital infrastructure, due to the lower number of inhabitants per hospital. This number alone is not enough to evaluate precisely the direct relationship between the number of hospitals per citizen and the emergency service capacity. Other numbers like the number of beds per hospital, the annual budget invested in public health care, the average age of the population, the list of common comorbidities of the population and, other environmental analyses are necessary to evaluate precisely the hospital services infrastructure.

5.2 What about the clusters?

The cluster model shows a strong relationship between the population size and the city area. The hospital infrastructure of each city was planned to handle emergency cases on an optimal condition, based on the average number of cases.

Analyzing the model, the cities of the bronze cluster (#3), except Kinshasa, would have a better quality of healthcare services, based on the lower number of citizens per hospital, the smaller population, and the area size.

Following this trend, the cities of the silver cluster (#2) have a bigger demographic density, raising the citizens per hospital rate. This proportional increase leads to a not so efficient healthcare service on pandemics and public calamity cases.

The cities of the gold cluster (#0) are counted among the biggest cities in the world, with heavy population traffic, intense demographic density and, a high number of citizens per hospital. Those cities probably would face hospital infrastructure collapse during a pandemic or a public calamity, due to the size of the population regarding the number of hospitals.

Finally, the platinum cluster is the extreme case, being Zhengzhou the biggest outlier. The city surpasses the population and area size of the second place in two and eight times, respectively. Beyond that, the city has the highest citizen per hospital rate, more than the double the second place. In a pandemic or public calamity, Zhengzhou certainly would face a very hard time taking care of all its citizens.

6. CONCLUSION

This analysis showed a strong relationship between the population size and the city area. Although, the population size is also related to the number of hospitals in a specific city. Besides that, the number of citizens per hospital probably has an important relationship with the healthcare services capacity and quality.

Cities with high demographic density and a high rate of citizens per hospital probably will face infrastructure issues during catastrophic events.

Due to the restricted number of features selected for this analysis, more research is necessary to evaluate precisely the correlation between different demographic, geographic, economic, environmental features and, the healthcare services capacity.