## Analyzing COVID-19 community transmission risk in Dakar's public venues.

## A. Introduction

### **A.1. Description of the problem and background**

### In December 2019, an outbreak of apparently viral pneumonia of unknown origin emerged in the city of Wuhan, in the Chinese province of Hubei and spread out in the entire world. This pneumonia is an infectious disease caused by a virus belonging to the coronavirus family, currently referred to as SARS-CoV-2 [1].

### As every country in the world at this current period, Senegal is facing the COVID-19 pandemic. Two types of transmission were identified: contact transmission (origin of the contamination is known) and community transmission (origin of the contamination is unknown)[2]. Community transmissions are more difficult to handle as investigations about the starting point may take time. At today's date, The Public Ministry of Health has reported 1,709 cases all over the country, where Dakar, the capital has registered the highest number of cases [3]. Among these cases, 184 are from community transmission.

### Dakar is a metro-pol with a high population density in a narrow area . As the incubation period is long enough , the others should avoid any contact / proximity to the focal point as much as possible. Thus, it is necessary to identify the vicinities of infected areas especially venues that are most frequented around this area where the density of population may yield to community transmission. In this project, we aim to identify public venues around infected sites and cluster them according the contamination risk.

### **A.2. Data**

To consider the problem, we use data from a public [GitHub repository](https://github.com/genova/covid19/tree/master/data) [4].The repository contains data from The Health Ministry which describe the evolution of the pandemic in Senegal. Note that The Health Ministry updates daily the data. The data used were lastly updated on *May 12th, 2020.* We identify as file of interest, zone\_c\_confirmesGeo.csv which depict the number of community cases per district . This file is structured as followed:

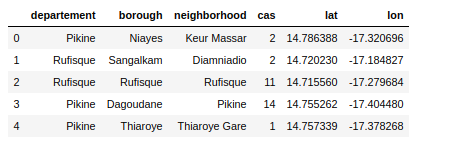
* district: name of district
* cas : number of community cases
* lat :district latitude
* lon: district longitude

We restrict the geographical area to Dakar using Google Maps API geocoding [5] . To stay consistent with the administrative division in Senegal [6], we describe data at a more fine-grained level. We set-up two additional attributes '*Borough*' and '*Department*' by collecting related information from Wikipedia and the Google Maps API geocoding . When values are not returned by the API (NaN values), they are set by default set to 'Dakar, Dakar'. All processed datasets are available in our GitHub Repository.

### **B. Methodology**

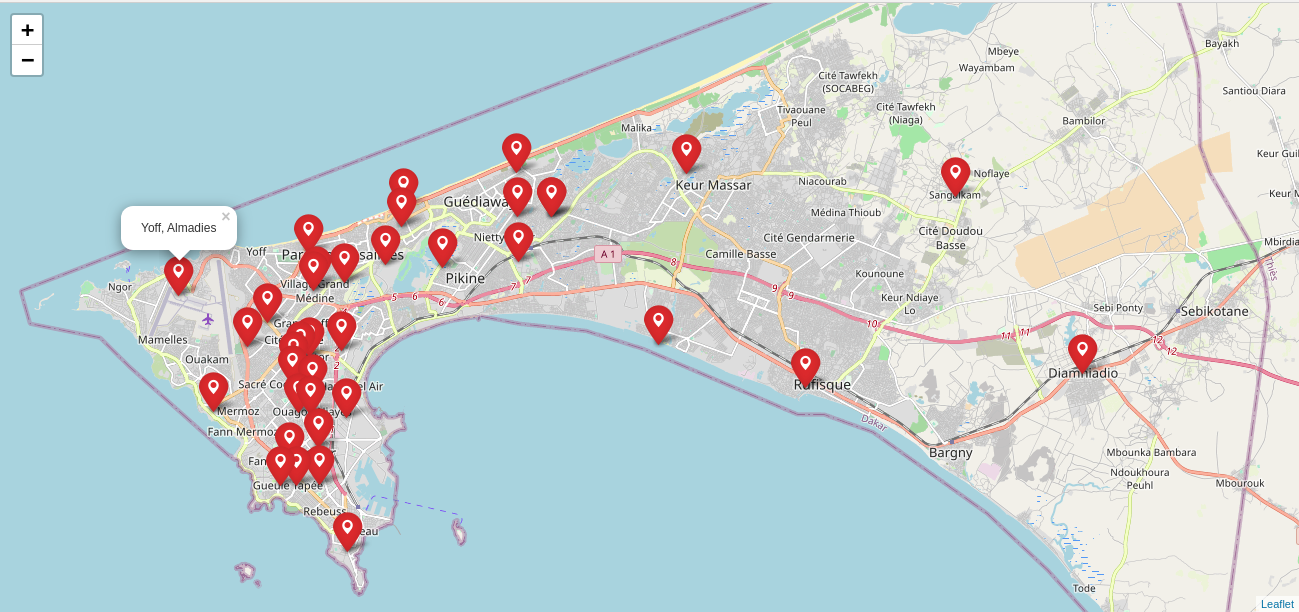
My master data has the main attributes *Department,Borough, Name , Numbers of cases, Latitude and Longitude* of places where community transmission has been detected in Dakar. It contains 36 rows. As mentioned, each update of the data set is *consider*.. , thus periodically the number of rows may change.

Ajouter une alternative



I used Python Folium library to visualize infected sites on a centered map of Dakar with dis superimposed on top and labeled with its name and borough. Coordinates of Dakar center are obtained using Google Maps API geocoding . I used latitude and longitude values to get the visual as below.

Ajouter une alternative



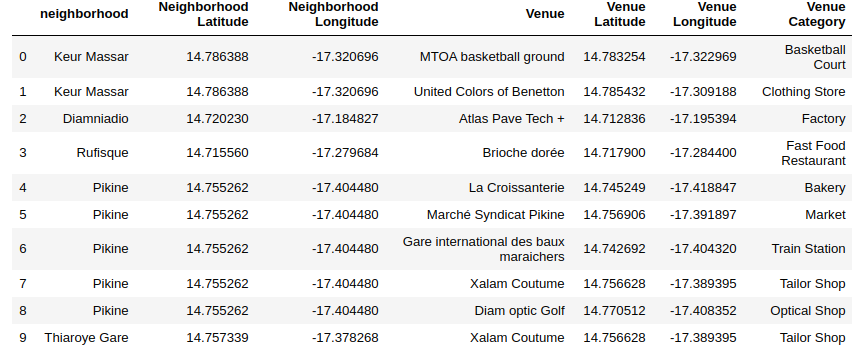
*Map of infected sites(* [*Link to interactive map*](https://ndiaye01.github.io/map_infectedsites.html)*)*

**B.1. Exploring neighborhoods**

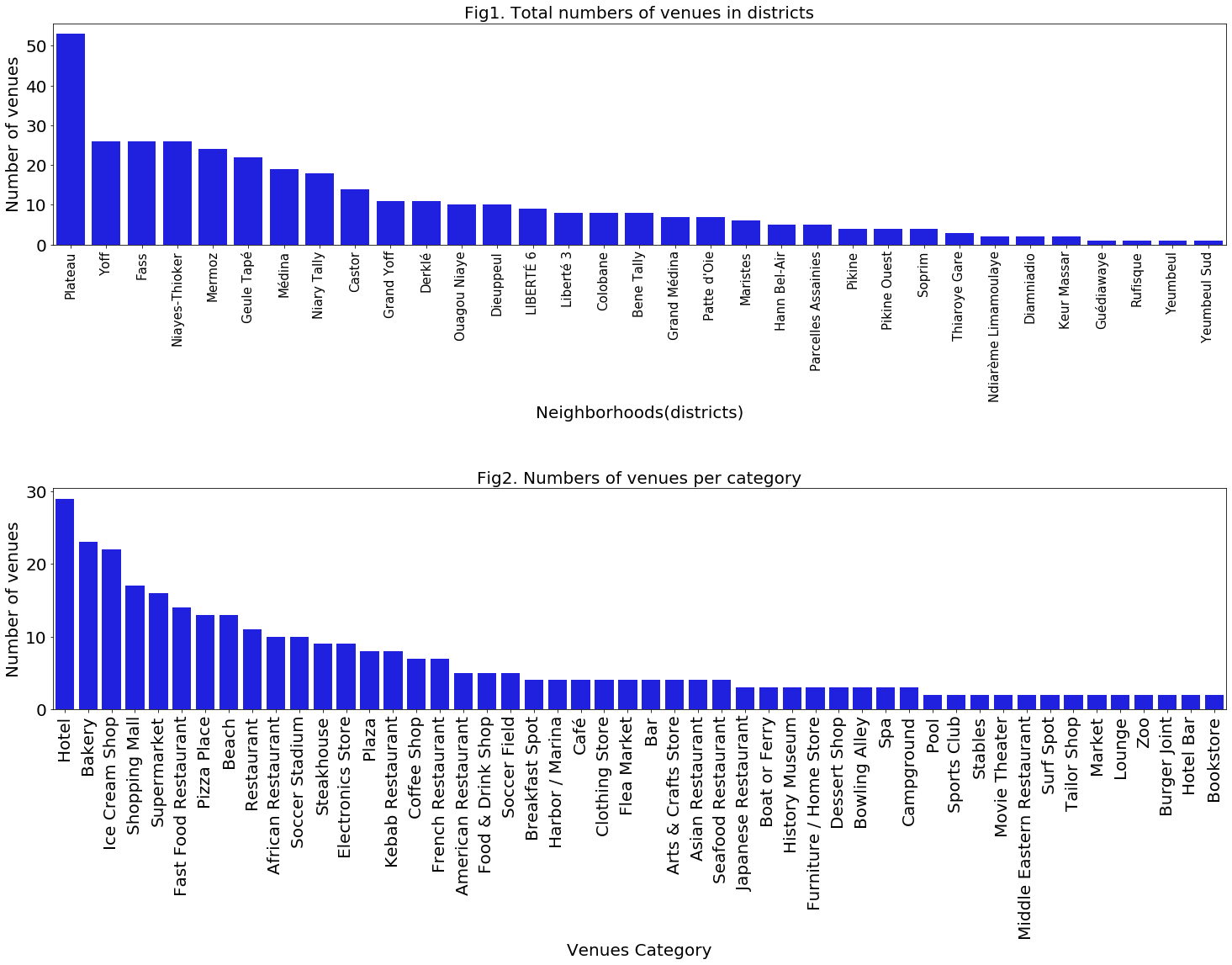
To explore around the infected sites, I utilized the Foursquare API [7]. I designed the limit as 100 venues and used different radius = 1000 meters for each site, from their given latitude and longitude informations. Thus, we are searching for dense venues that are 1km around each infected site.

Here is a head of the resulting dataframe with venues name, category, latitude and longitude informations from Foursquare API.

Ajouter une alternative



In summary, 355 venues from 78 unique categories were returned by Foursquare. Let's check how many venues were returned for each neighborhood (district) .Ajouter une alternative

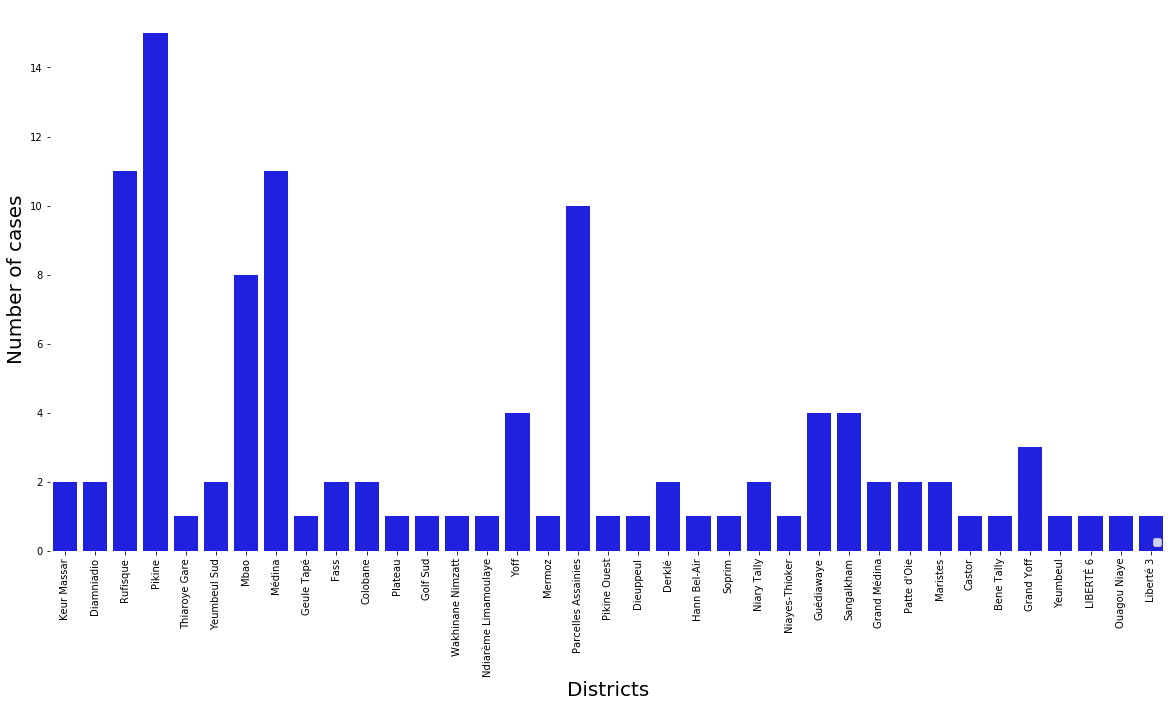


We can see in Figure 1 that the district Plateau reached the 100 limit of nearby venues.\*\*\* This is the city center. On the other hand, most of the districts are below 10 nearby venues. Three districts have zero nearby venues: Mbao, Wakhinane Nimzatt, and Sangalkam. These are unpopulated districts. As shown below in figure 2, most of venues are Hotels, Ice cream shops and bakeries.

However, the result doesn’t mean that inquiry run all the possible results in neighborhoods (districts). Actually, it depends on given the radius information and we just run a single value for the radius . We can increase the possibilities with Neighborhood informations by varying the radius value.

B.2 Clustering infected sites and its nearby venues

To cluster infected sites, I used unsupervised learning K-means algorithm. K-Means algorithm is one of the most common cluster method of unsupervised learning. Intuitively, based on the context of our problem, a contamination risk level can either be High, Moderate or Low. Let's take a look on data, to ensure that 3 is the optimum value for the K-Means number of clusters.

Ajouter une alternative

When we examine above graph,we can easily group districts as followed: high ,moderate and low number of cases.

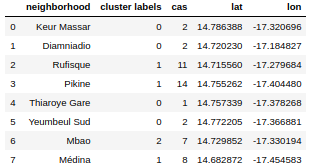
I will run K-Means to cluster the infected sites into 3 groups using the dataset features cases, latitude and longitude. We further create a clustered map of their vicinities. The initialization method of the centroids is 'k-means++' to speed up convergence.

C. Results

The above table shows the labels assigned to each infected site. As it seems in below table, we can define clusters as below:

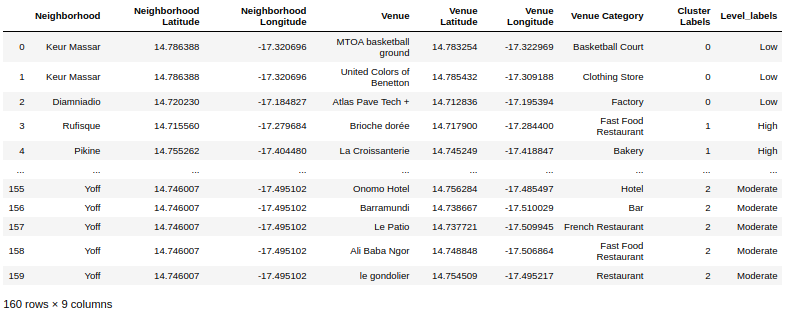
* Cluster 0 : “Low Level Risk”
* Cluster 1 : “High Level Risk”
* Cluster 2: “Medium Level Risk”

Ajouter une alternative



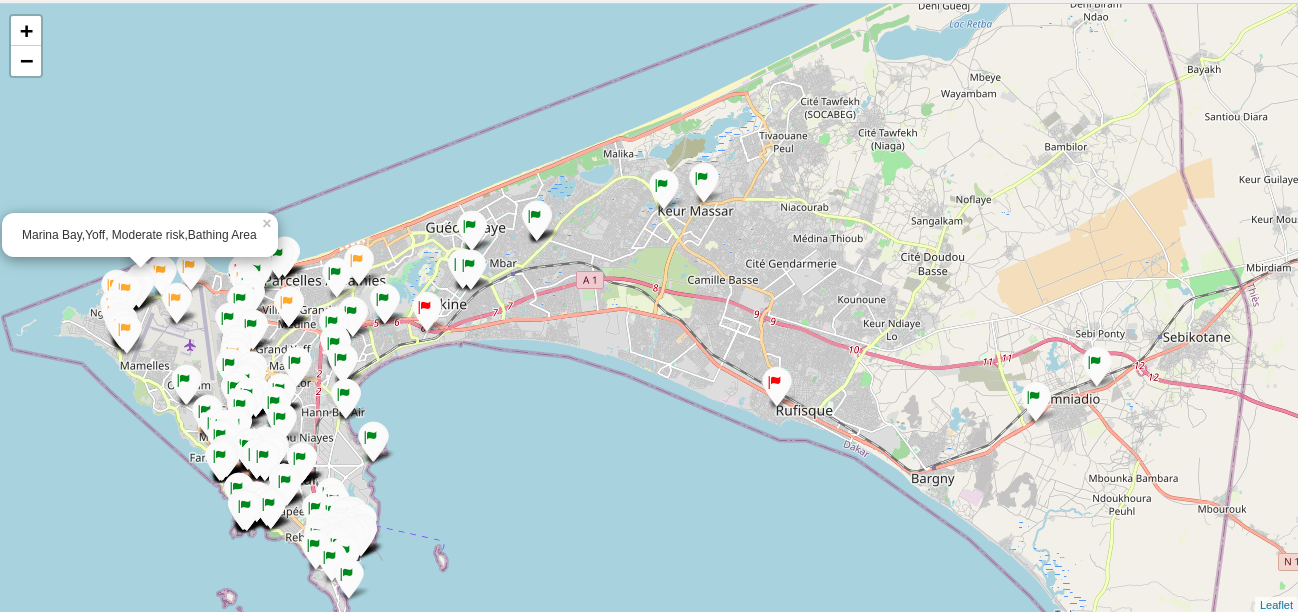
We merge the above table with the vicinities table with venues name, category, latitude and longitude informations. Thus, to each venue is assigned its district's label. You can now see labels and levels columns as the last two ones in below table, for each venue.

Ajouter une alternative



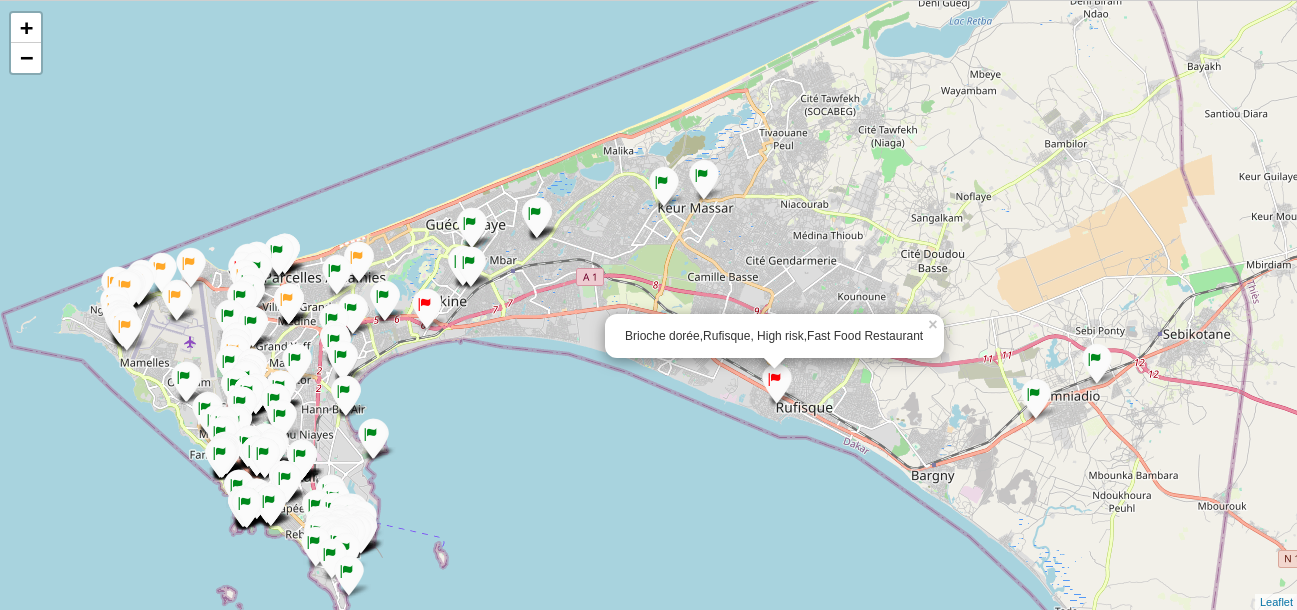
Finally, I created the clustered map of venues in the below. Each venue is labelled with its name, borough, contamination risk level and category. Regarding color assignment, high risk level venues are marked with a red flag, Moderate 's one with an orange flag and low's one in yellow.

Ajouter une alternative



*Link to the interactive map:* [*https://ndiaye01.github.io/map\_clusters\_2000.html*](https://ndiaye01.github.io/map_clusters_2000.html)

Ajouter une alternative



*Link to interactive map:*[*https://ndiaye01.github.io/map\_clusters\_2000.html*](https://ndiaye01.github.io/map_clusters_2000.html)

D. Discussion

Our results allow us to identify that the center and the south of Dakar which are the densest areas are relatively low- exposed. The venues in West of Dakar such as the Yoff Beach and its Marina Bay tend to be moderately exposed to community transmission. The highly exposed venues are located in the suburbs (Pikine, Rufisque ,Mbao). But the exploration shows these areas are not dense. The results are in range with the contacts transmission maps, which identify the similar areas of concern.

Our model updated the data daily, which provides visibility in real time. However, our results depend systematically on the radius as it can increase the possibilities with neighborhood informations. For example, for radius= 1000, 139 venues were found . T[he resulting map](https://ndiaye01.github.io/map_clusters_1000.html) is still consistent with observations made above.

However, the fact that the districts are squeezed, causes the city to have a very intertwined structure. As there is such a complexity, As very different radius can be tried in exploration, this lead us to certain observation. Not every radius value can yield the same high quality results for this metropol as high radius may create overlapping zones. We experimented it for radius= 5000( [cf. interactive map](https://ndiaye01.github.io/map_clusters_5000.html)).

### **F. Conclusion**

In this study, we identify which public places in Dakar are exposed to COVID-19 community transmission and to what extent , given a radius. We compared three level of risk transmission: low , high and moderate. After examining the clustered map, our statistical analysis concluded that the overall risk of transmission in Dakar's venues is relatively quite moderate, because where the population is dense is low-exposed and besides the highly-exposed areas are sparse. Of course, theses observations may change daily according to the pandemic evolution.

In community transmissions, one will trace the moves of the individual as well as his interactions, particularly in his nearest neighborhood. Rating the risk of transmission of a public venue could be a precious information especially for directing transmission surveys in areas of high density such as Dakar. Similar data analysis types could be useful in certain decision-making such as the re-opening of public places which is controversial today within the country. Depending on their proximity, with the infectious site, some may be accessible again to the public else not.

In future studies, we figure out

### **G. References**

* [1] [Institut Pasteur](https://www.pasteur.fr/en/medical-center/disease-sheets/covid-19-disease-novel-coronavirus)
* [2][C'est quoi la transmission communautaire?](https://www.seneplus.com/sante/le-pr-daouda-ndiaye-explique-le-processus)
* [3] [Situation of Covid-19 in Senegal](http://www.sante.gouv.sn/)
* [4] [Public GitHub Repository](https://github.com/senegalouvert/COVID-19/tree/master/data)
* [5] [Google Map](https://www.google.com/maps/)
* [6] [Dakar -Wikipedia](https://en.wikipedia.org/wiki/Dakar#Geography)
* [7] [Foursquare AP](https://developer.foursquare.com/)I