THE “15 MINUTES CITY”

A DATA SCIENCE PROJECT

JULY 2020

## 1. INTRODUCTION

### 1.2 BACKGROUND

In June 2020 mayoral elections took place in Paris, France. The winning candidate wants to make Paris a city where anyone living in the city can reach key services and amenities within a 15 minutes walk or bike ride from home. Achieving this vision would yield several benefits, among which: improving quality of life, reconnecting people with their neighborhoods and improving air quality by reducing drastically the need to drive.

More information about this vision can be found in this article in English:
https://www.fastcompany.com/90456312/pariss-mayor-has-a-dream-for-a-15-minute-city


### 1.2 CHALLENGE

Such an ambitious vision for a large city could be complex to implement and can generate skepticism.

Depending on the type of services or amenities, the approach to achieve this goal and the current gaps to address would vary.

For example, if in a given neighborhood many homes are not within the expected travel time from a doctor's office, some of the challenges could be that there are no appropriate buildings that can accommodate a medical practice, healthcare professionals might have to be incentivized to practice in this area, other costs could be required to enable 15 minutes access to medical services... Basically, multiple types of challenges could come up: financial, logistical, economical, demographic...

## 1.3 QUESTIONS THAT WE WILL TRY TO ANSWER

As we just explained, the approach and cost of enabling the "15 minutes" vision can vary depending on the neighborhood and the type of services that the city of Paris hopes to make available within a 15 minutes walk bike ride from anyone's home. Therefore, our goal for this project will be to answer the following questions:

1. **What is the overall current state of readiness of the city and the districts within it in respect to this "15 minutes" goal?**

2. **Can we identify neighborhoods that share similar characteristics and group them together so that specific dedicated action plans can be defined to enable the "15 minutes" vision in these neighborhoods?**

***Note:*** *Paris is divided into 20 districts. The district of an address in Paris can be determined based on the postal code. The last 2 digits of the postal code indicate the district. Example: postal code 750<u>18</u> is for the 18th district.*

**IMPORTANT NOTE: for this project we will only focus on <u>walking</u> distances and ignore bike rides.**

## 1.4 WHY DO THESE QUESTIONS MATTER?

Conducting this analysis and answering these questions would enable the city of Paris and the districts within in to come up with specific strategies and identify potential readiness costs associated with groups of related neighborhoods. This is an important objective because it will enable better planning of this initiative from a financial and tactical perspectives, and could potentially provide useful insights for other cities looking to achieve similar objectives. For example, the city of Ottawa in Canada is currently investigating the same goal (*source: article referenced above*).

## 2. REQUIRED DATA TO ENABLE THIS PROJECT

To answer our two questions, we will use datasets generated from Foursquare and from the City of Paris Open Data website, and create additional datasets derived from the first two. First, let’s review a few assumptions and decisions, before reviewing our data choices in detail.

## 2.1 IMPORTANT! DECISIONS, SIMPLIFICATIONS AND ASSUMPTIONS FOR THIS PROJECT

In order to complete this project, we are making a few assumptions and adopting a few simplifications described below:

1. Despite several attempts Foursquare was not returning proper datasets of residences located in Paris. Data were missing and other type of venues kept being included despite using relevant residence category IDs provided by Foursquare’s online help. Instead we will use a dataset publicly available via the Paris city Open Data website, this dataset only includes social housing buildings managed by the city.
   Our approach for this project will be generic enough to accommodate other sources of residential buildings, so if a more complete source was available it would be minimal work to adjust the initial dataset construction. We would only have to modify the code retrieving data from the data provider.

2. We will measure the distance between residential buildings and amenities using a straight line. This is of course not realistic as in a city it is often impossible to follow a straight line to go from one point to another. Calculating the true distance would require to use some data from online services that provide travel time calculations (e.g

Google Maps). This could be the purpose of a future revision of this project.

3. We will use a walking speed of 5km/h. This speed would of course vary depending on the person's health and fitness and on the terrain.

4. We will limit our analysis to a few types of amenities and services, and to a few districts in Paris, in order to work with a reasonable dataset and minimize processing times. This is also largely due to some limitations in number of records that can be retrieved from Foursquare.
   - Services to be included: a few categories of medical services (described in section 2.2 below)
   - Districts: $17^{th}$, $18^{th}$ and $19^{th}$ districts, located in the northern part of Paris

   However our data design approach and project methodology will be generic enough to include more services or districts. This decision only impacts the number of records included in our datasets.

## 2.2 SOURCE DATASET 1: SERVICES

We select the type of services for which we want to calculate a travel time from residential buildings.

| | |
|---|---|
| **Data source** | Foursquare venues, limited to some sub-categories specified below |
| **Target data storage** | A panda dataframe called **med_services_clean** |

As explained above, we will only include certain categories of Medical venues for this project. The list of Foursquare venue categories with their category ID is available at this location:

https://developer.foursquare.com/docs/resources/categories

Examples of venue categories and associated category IDs to be included in this project:

| **Key medical services** | |
|---|---|
| Doctor's office | 4bf58dd8d48988d177941735 |
| Dentist's office | 4bf58dd8d48988d178941735 |
| Hospital | 4bf58dd8d48988d196941735 |

We will also need location information (i.e geographical coordinates) for all the services that we have decided to include in our analysis.

This dataset will be populated using the following criteria:

| |
|---|
| venues only |
| venue categories in the list of venue category IDs listed above |
| country = France |
| postal code starting with 75 because French addresses located in Paris all have a postal code starting with |

| |
|---|
| 75 (addresses outside of Paris do not have a postal code starting with 75) |

**Structure of med_services_clean:**

| COLUMN | SOURCE / DEFINITION |
|---|---|
| ServiceID | Foursquare Venue ID |
| ServCat | Foursquare Venue Category |
| ServLat | Foursquare Venue Latitude |
| ServLong | Foursquare Venue Longitude |
| ServPostCode | Foursquare Venue Postal Code |

## 2.3 SOURCE DATASET 2: RESIDENTIAL BUILDINGS

We need to obtain location information of all residential buildings in the districts we have decided to analyze ($17^{th}$, $18^{th}$ and $19^{th}$ districts as explained above) so that we can calculate their minimal distance to amenities included in our analysis.

| | |
|---|---|
| **Data source** | JSON dataset of social housing residences provided by the Paris City Open Data project. |
| **Target data storage** | A panda dataframe called **residences_df** |

We will also need location information (i.e geographical coordinates) for all the residences that we have decided to include in our analysis.

This dataset will be populated using the following criteria:

| |
|---|
| postal code = 75017 or 75018 or 75019 to only include addresses in the $17^{th}$, $18^{th}$ and $19^{th}$ districts of Paris |

Later, in a separate dataset, for each residential building we will calculate a boolean flag indicating whether a residential building is meeting the 15 minutes goal for ALL services included in this project

**Structure of residences_df:**

| COLUMN | SOURCE / DEFINITION | CLARIFICATIONS |
|---|---|---|
| ResID | Foursquare Venue ID | |
| ResLat | Foursquare Venue Latitude | |
| ResLong | Foursquare Venue Longitude | |
| ResPostCode | Foursquare Venue Postal Code | |
| ResDist | = Last 2 digits of ResPostCode | CALCULATED FEATURE |
| 15MinStatus | Boolean value<br>= Boolean multiplication of 15MinFlag values | CALCULATED FEATURE<br>Use 15MinFlag for each type of service calculated in a |

| | | |
|---|---|---|
| | associated to a ResID | separate dataframe described below |

## 2.4 DERIVED DATASET: RESIDENCE ASSESSMENT BY SERVICE CATEGORY

Using our 2 source datasets and some calculations we will create a third dataset, stored in a panda dataframe called **resid_med_serv_df**.

Once this new dataframe is fully populated with the calculated features flagged below we will be able to run some analysis on the 15 minutes readiness overall, by district, by type of service. We will also be able to do some analysis on travel times. This dataframe will also enable some data visualization using a map of the city.

This dataframe will also serve as a base to create a transformed dataframe suitable for clustering.

**Structure of resid_med_serv_df:**

| COLUMN | SOURCE / DEFINITION | CLARIFICATIONS |
|---|---|---|
| ResID | ResID from residences_df dataframe | |
| ResLat | ResLat from residences_df dataframe | |
| ResLong | ResLong from residences_df dataframe | |
| ResDist | ResDist from residences_df dataframe | |
| ServCat | ServCat from med_services_clean dataframe | |
| MinTravelTime | = Minimum of all travel times from a given ResID to all Services for a given ServCat | CALCULATED FEATURE<br>We will create a function that calculates the travel time from a given Residence to all Services, for a given ServCategory, the function will return the minimum travel time value identified for a given ServCategory. This function will use the geographical coordinates of Residences and Services |
| 15MinFlag | = 1 if MinTravelTime ≤ 15<br>= 0 otherwise | CALCULATED FEATURE |

# 3. METHODOLOGY

**Step 1 – Import the required data**

As described above, from 2 sources: Foursquare for medical service venues, the Paris City open data project for social residences.

**Step 2 – Data preparation**

We limit our social residences dataset to the 17$^{th}$, 18$^{th}$ and 19$^{th}$ districts of Paris. This will help limit the volume of medical service venues pulled from Foursquare and comply with the limitations in terms of data volume that can be retrieved with a basic Foursquare account.

We limit the type of medical service venues included in the dataset pulled from Foursquare to limit the volume of data to be processed and also because some of these services are not relevant (they are not true medical services).

**IMPORTANT:** We include medical services from all districts in Paris in case the closest service to a residence is in another district, not necessarily in the 3 districts where our residences are located.

We also make sure that we do not have any duplicate medical service venues. We retrieve them by looking within a large radius around each residence so it is possible that some medical service venues are retrieved multiple times. A comparison of the medical service venue dataframe shape before and after deleting duplicate records confirms this assumption.

**Step 3 – Calculate minimal walking distances from each residence to each category of medical service venue**

This step enables us to create a dataframe that will capture for each residence the minimal walking time to each category of medical service and whether this time is under 15 minutes or not (boolean flag).

This dataframe includes additional information that will be useful later for our analysis and map plotting (residence ID, postal code of the residence from which we derive the district, residence latitude, residence longitude).

Later during our exploratory data analysis we will derive another dataframe from this one to obtain for each residence an aggregated status indicating whether a residence is under 15 minutes of ALL medical services or not.

**Step 4 – Exploratory data analysis**

This analysis involves 2 approaches:

- Some basic statistical analysis illustrated by graphs
- Some plotting of visual indicators on a map of Paris

Statistical analysis conducted:

- Check for each medical service category the number of residences meeting the 15 minutes walk target
- Distribution of travel times across all medical service categories, with mean and median
- Distribution of travel times by medical service category, with mean and median
- Ratio of residences meeting the 15 minutes target for all services, by district
- Ratio of residences meeting the 15 minutes target by medical service, by district

Map visualizations:

- Identify residences that meet or do not meet the 15 minutes criteria overall
- Identify number of distinct medical services within 15 minutes walk for each residence

Throughout this phase we will generate some new dataframes based on our 2 original residences and medical services dataframes in order to organize data in a convenient way for the analysis that we plan to conduct.

**Step 5 – Clustering & clusters interpretation**

Our goal with clustering is to attempt to group residences that have common characteristics in terms of 15 minutes access to medical services.

Interpreting the clusters, by confirming that they have a somewhat identifiable logic explaining why some residences belong to the same cluster, could help determine a specific approach to address gaps in terms of access to medical services.

The clustering approach will be based on a K Means approach.

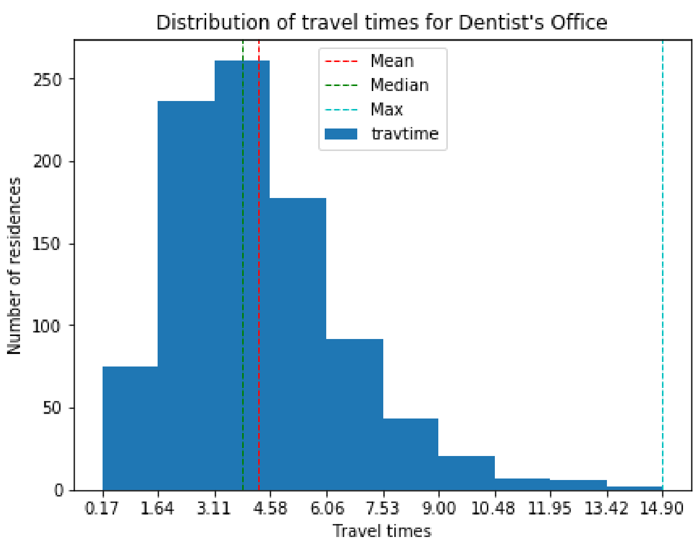
Our clustering features will be the boolean statuses describing for each medical service category whether a given residence is under 15 minutes walking distance or not.

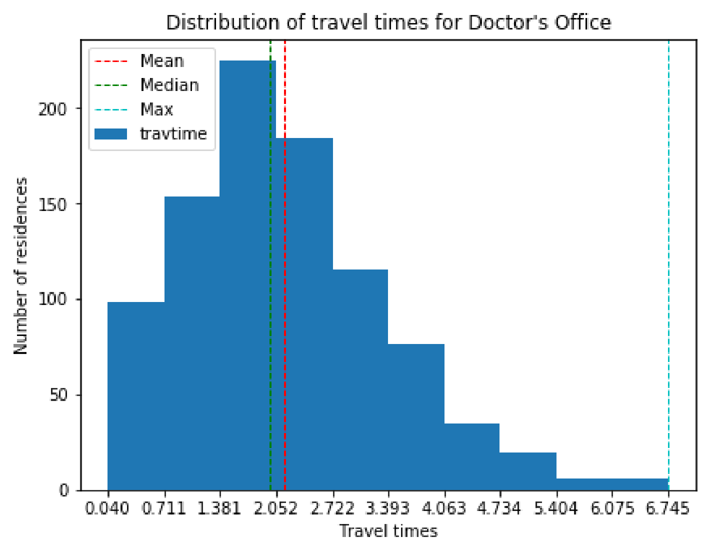
## 4. RESULTS & DISCUSSION

### 4.1 Exploratory data analysis

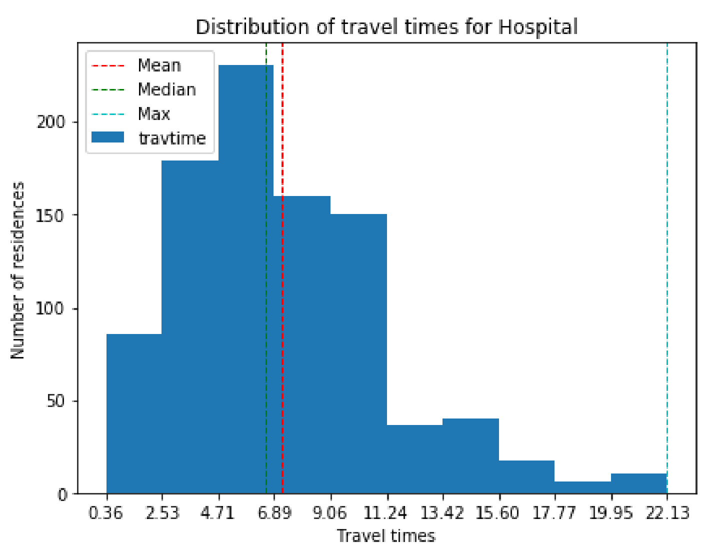
The main outputs of our exploratory data analysis show that:

- All residences in all 3 districts are within a 15 minutes walk from a Doctor's Office and a Dentist's Office.
  This is demonstrated by these 2 histograms of travel times to these 2 medical services, where the max value is under 15 minutes.

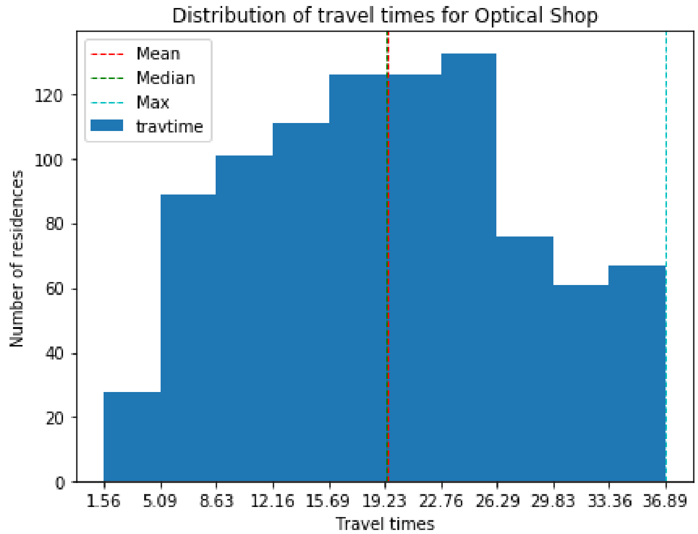


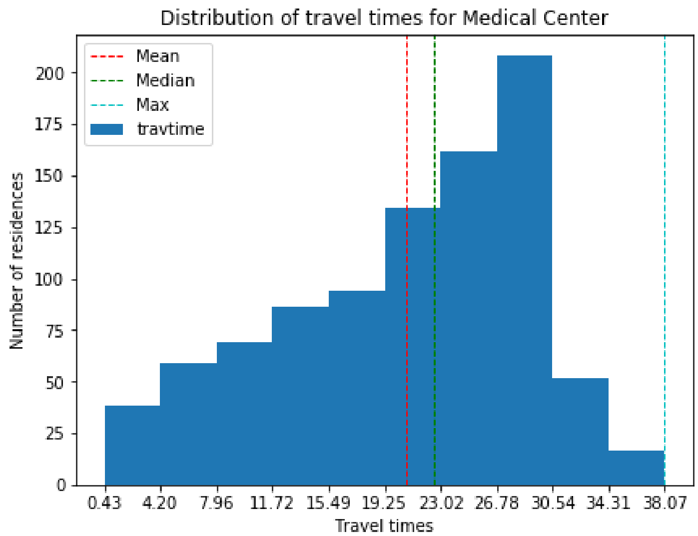




A similar type of histogram shows that a majority of residences are within 15 minutes of a hospital. This can be inferred from the position of the median and the height of the bars above 15 minutes.

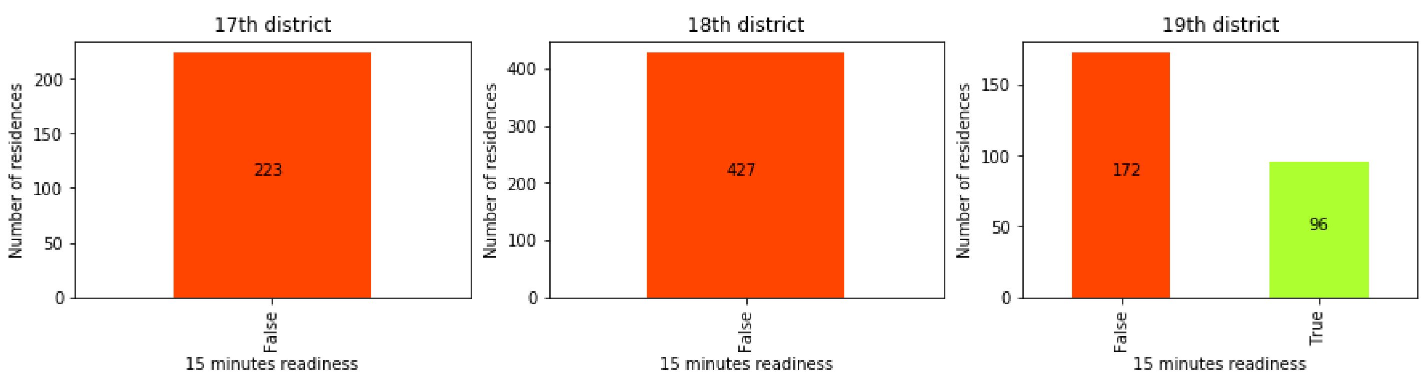


For optical shops and medical centers this is not the case:

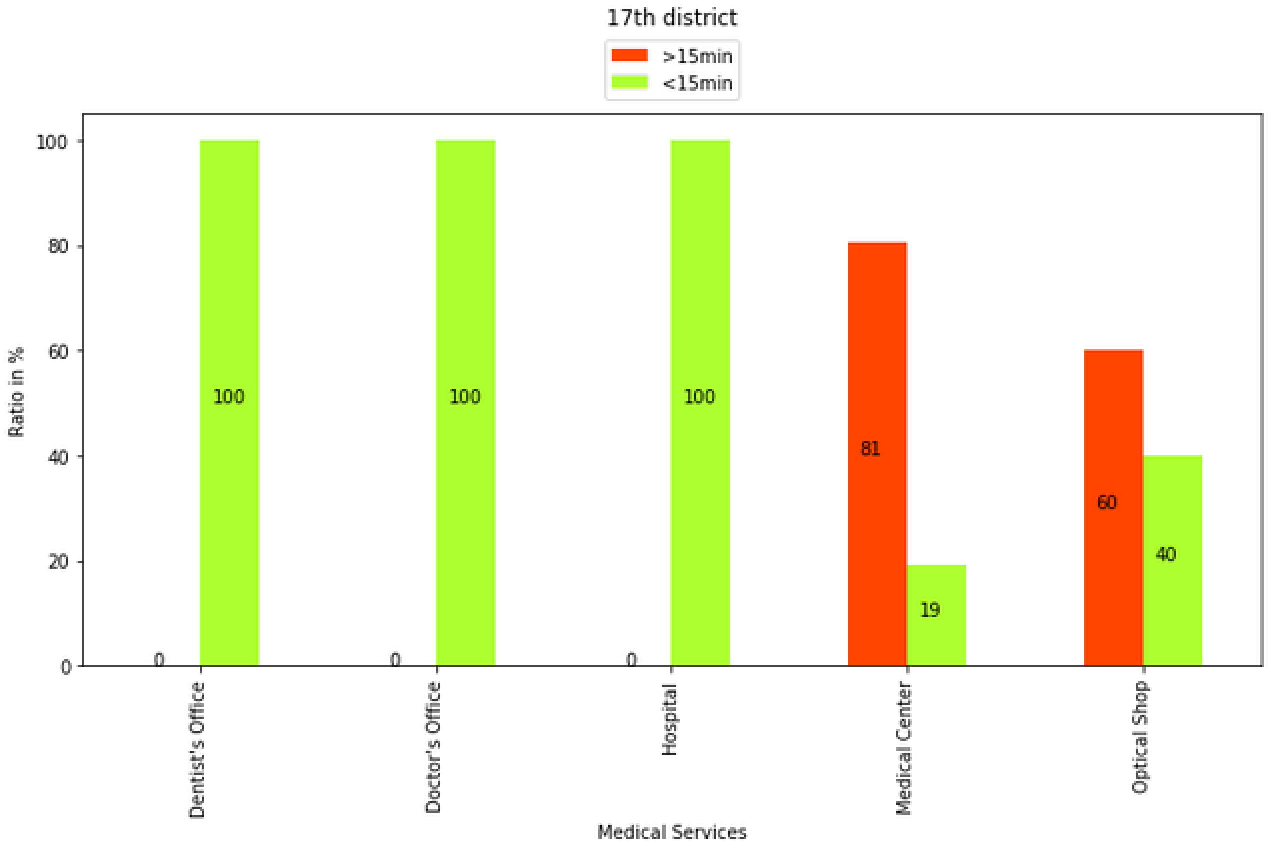


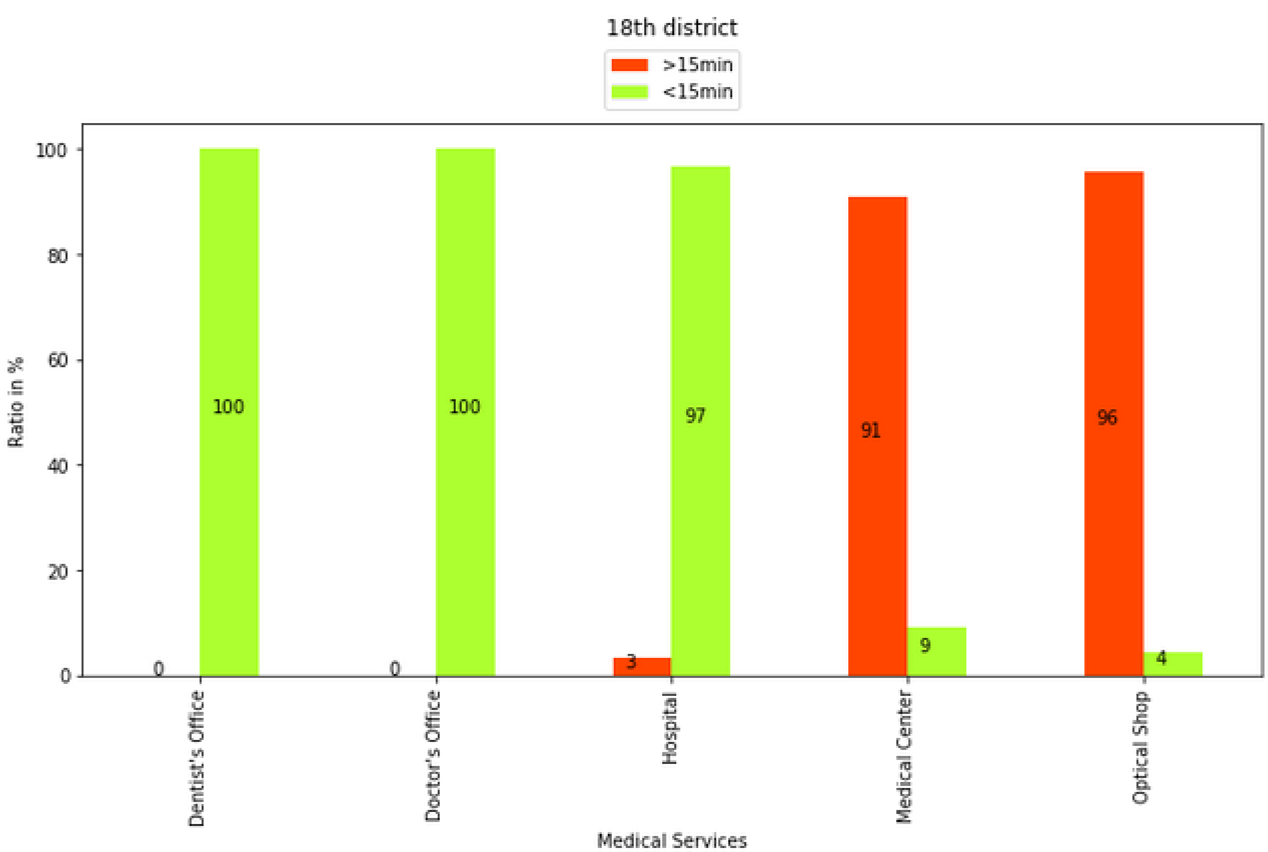

- Only in the 19th district do we find residences (96) that are within a 15 minutes walk of **all** medical service categories considered in this project.



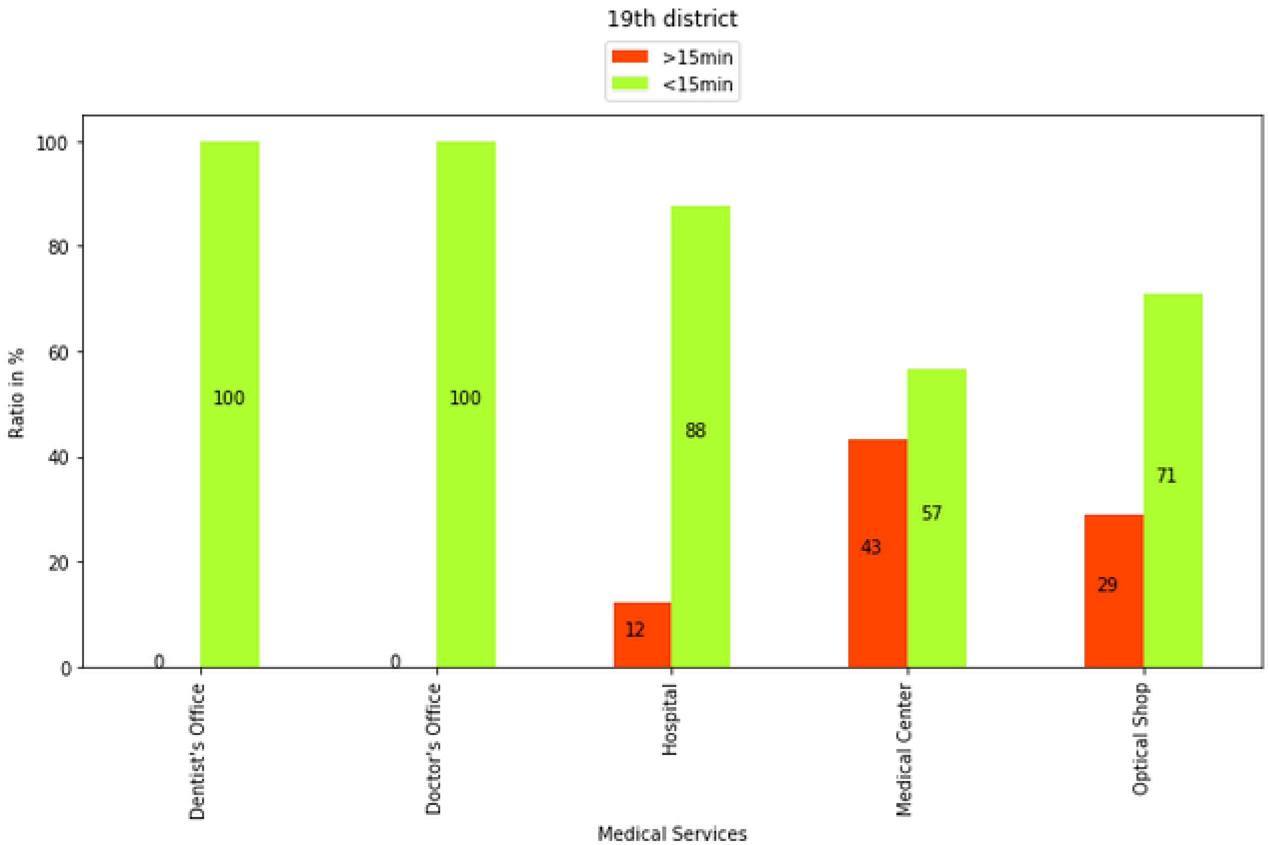

- Detailed analysis by district
    - In the 17th district all residences are within 15 minutes of a hospital, but a majority are NOT within 15 minutes of a medical center (81%) or an optical shop (60%).




- In the 18th district, a large majority of residences are within 15 minutes of a hospital (97%), but a large majority are NOT within 15 minutes of a medical center (91%) or an optical shop (96%).



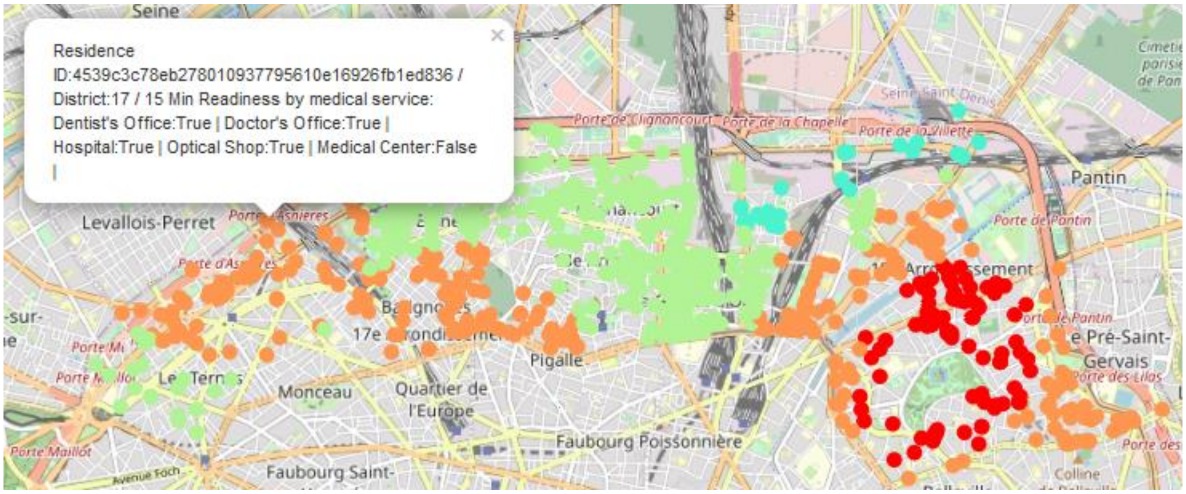
- In the 19th district, a large majority of residences are within 15 minutes of a hospital (88%) and a majority of residences are within 15 minutes of a medical center (57%) or an optical shop (71%).




In summary, for the most common medical needs (doctor's office, i.e general practitioners, and dentists) the goal of having a reasonable walking trip from home is achieved. For hospitals the situation is overall positive and a small fraction of residences are not within 15 minutes of a hospital. However, for other types of medical services there needs to be significant improvements in terms of short access.

Overall the 19th district is the one with the best access overall to medical services with a majority of residences being less than 15 minutes away from all medical services.

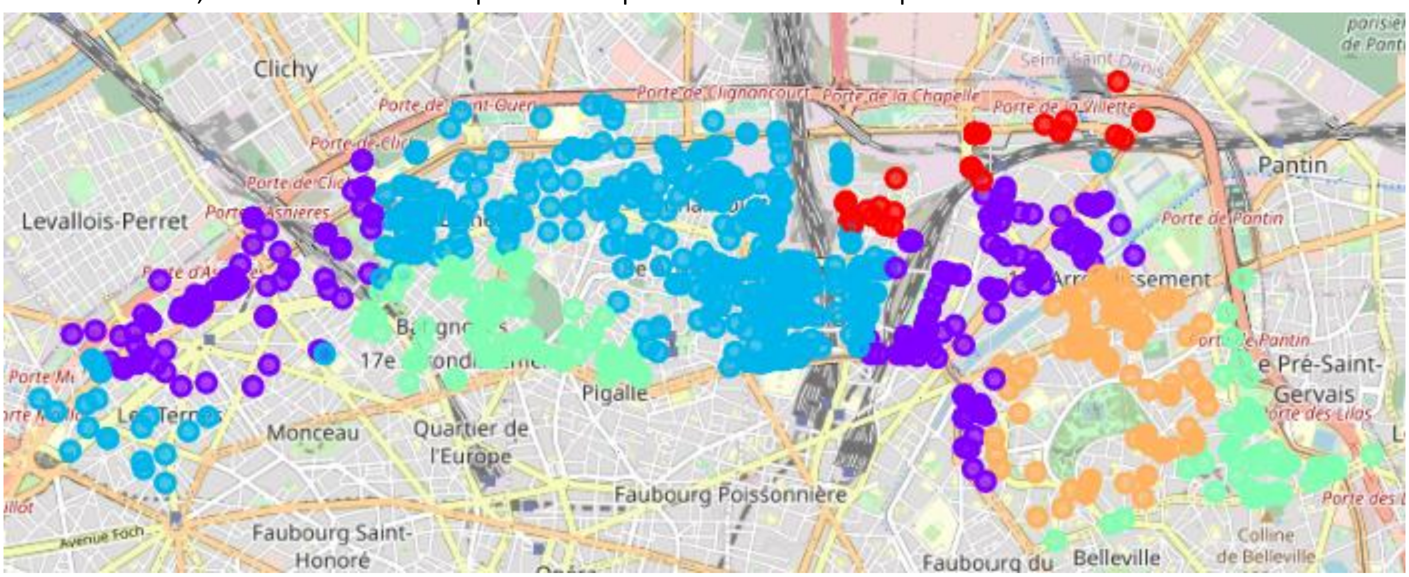
In addition, a clickable map in the Jupyter notebook enables to visualize residences and the number and type of medical services within a 15 minutes walk.



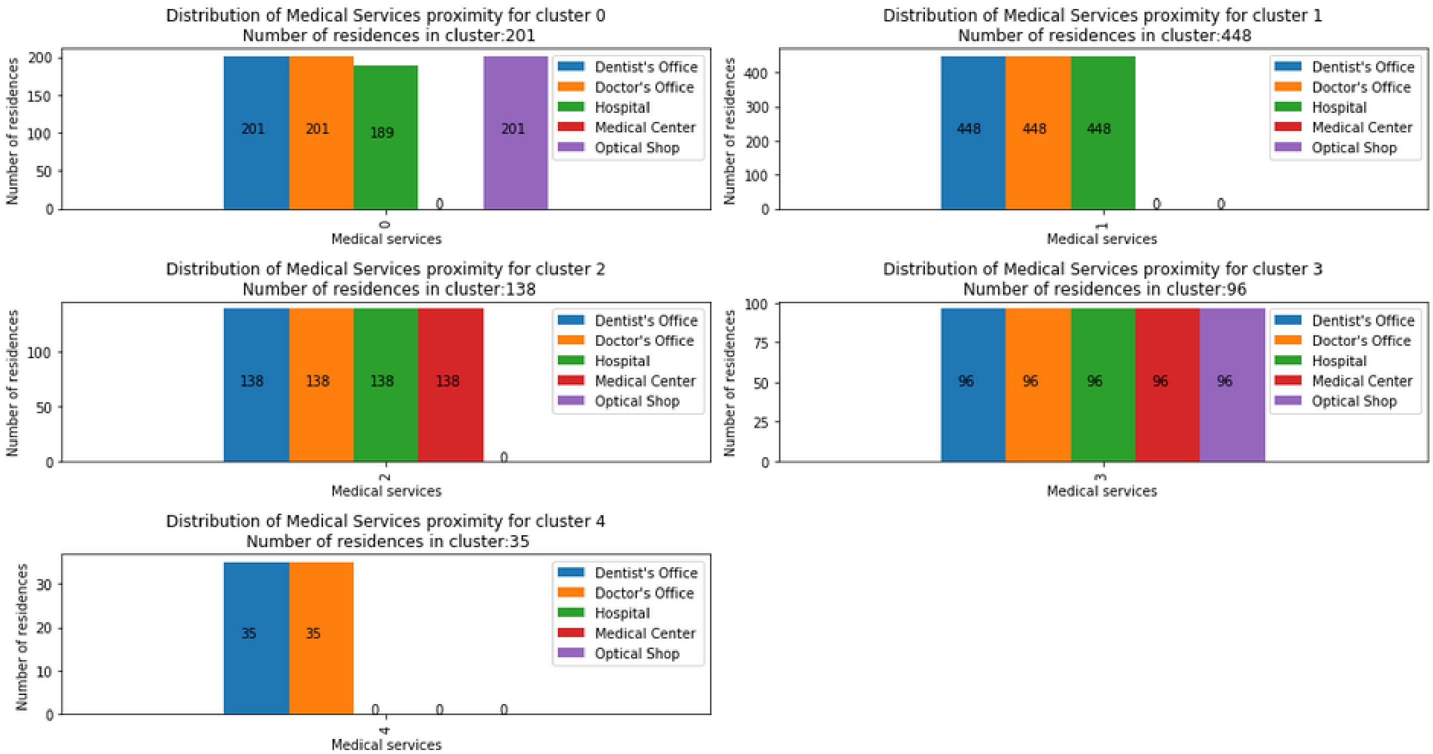

## 4.2 Clustering outcome

The end result of the clustering process applied to a map of Paris shows that:

- Clusters tend to group residences that are in continuous areas, i.e we do not see much dispersion in the way residences within the same cluster are plotted on the map. They tend to form continuous areas that are separate from each other, even if 2 clusters are split into 2 separate areas on the map.




An analysis of the distribution of medical service 15 minutes flags within each cluster show some clear patterns explaining why some residences belong to the same clusters.




Residences within the same cluster share common properties:

- **Cluster 0 (purple on the map)** contains residences that are generally within 15 minutes walking distance of all medical services except medical centers.
    - **NOTE:** we notice however that a small number of residences (12) in this cluster are not within 15 minutes of a hospital, but this number is small enough to consider a consistent approach for all residences in this cluster. The 12 exceptions can be investigated on a case by case basis.

- **Cluster 1 (blue on the map)** contains residences that are within 15 minutes walking distance of all medical services except medical centers and optical shops.

- **Cluster 2 (green on the map)** contains residences that are within 15 minutes walking distance of all medical services except optical shops.

- **Cluster 3 (orange on the map)** contains residences that are within 15 minutes walking distance of all medical services.

- **Cluster 4 (red on the map)** contains residences that are within 15 minutes walking distance only of doctor's offices and dentist's offices.

Having clear patterns defining these clusters and the fact that these clusters are clearly separated from each other on the map can simplify the approach of addressing the 15 minutes gaps in these clusters and can enable the definition of strategies specific to each cluster.

## 5. CONCLUSION

Our goal for this project, outlined earlier in this report, was to address the following questions pertaining to easy access to medical services:

- What is the overall current state of readiness of the city and the districts within it in respect to this "15 minutes" goal?
- Can we identify neighborhoods that share similar characteristics and group them together so that specific dedicated action plans can be defined to enable the "15 minutes" vision in these neighborhoods?

We have been able to answer both questions:

- For the districts that we selected for this project, we have been able to determine the "15 minutes" readiness of these districts, at the residence level, by type of medical service. We obtained some numerical results and were also able to visualize on a map these results. As a result:
  - we could determine “15 minutes” readiness overall and by residence, and by medical service
  - we identified some differences of readiness that can help determine appropriate action plans.
- We also identified common readiness characteristics shared by residences and even neighborhoods within the districts via a clustering exercise, supplemented by a map visualization of the clusters.

This approach can benefit city planners looking to identify and address “15 minutes” gaps in the city, and devise local strategies specific to a district or to a neighborhood within a district. The readiness statistics obtained can also be used as a baseline to track the progress and effectiveness of measures introduced to address these gaps.

**Beyond this project**

Assuming no limitations in terms of data access and API costs, without changing the overall methodology and the goals of this project, and with minimal code changes, we could definitely introduce some improvements in terms of data included and walking time calculations:

- Assuming we did not have any of the data acquisition limitations exposed above (limited volume for free accounts, inability to access a list of all residential buildings/homes and not just social housing) we could expand our analysis to the whole city and all types of homes using the same approach. We could also expand it to other types of services, not just medical. The code was written in a generic way to allow this scalability.
- Another improvement would be to use a true walking time calculation method. We would just have to modify a portion of code within the dedicated function that we created for this purpose. The function’s arguments and outputs would remain the same.

However, this project confirmed that with the simple approach we used we can confirm the feasibility of evaluating a city’s readiness for a “15 minutes” goal as defined in the introduction of this report, and obtain some granular evaluations by district/neighborhood and by service category.