

15'min Emergency

IBM Capstone Project

1. INTRODUCTION:

The aim of this project is to test the readiness of the city amenities where each residence can reach any necessary amenities like medical services in an emergency within the 15 minutes of timespan. This project can be further expanded to other amenities which can help in better city planning, resulting several benefits such as improved life quality, near reach of amenities causing less driving, controlling traffic and air pollution.

The primary approach in implementing such as city-wide project, depends on the neighborhood and the essential service city wants to enable the '15min Emergency'. Here in this experiment, we will be focusing on the Downtown neighborhood of the Washington DC for the medical services available in the reach from downtown.

With the help of this experiment, we will try to answer some of the questions like:

- 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?

2. DATA COLLECTION:

For the experiment, we will data of the each or maximum possible residence of building the desirable city or neighborhood. In this project, we will be focusing on Washington DC specifically Downtown of the city. We will the coordinates of residence to locate the spot and find the medical service in the desirable radius. For this data, we will use the building benchmark data sourced from Open Data DC which contains all the required information. We will use the following data from the dataset:

- PM_PROPERTY_ID: A unique key given to all residence
- LATITUDE: The latitude of residence
- LONGITUDE: The longitude of residence
- PM_PROPERTY_NAME: The operational name of the residence/building
- ADDRESS: Address of the property
- POSTAL_CODE: The postal code of the neighborhood

For the next part of the project, we need to search for all the medical service/categories available in the radius of every residence of the downtown using the coordinates from the above dataset. For this, we will use Foursquare API to search for the medical services in reach. We will use the following data from the dataset:

- Venue Category: The category of the medical service
- Venue Category id: The category id of the medical service
- Venue Id: The unique id of every medical service
- Venue Latitude: The latitude of the venue
- Venue Longitude: The longitude of the venue
- Venue Name: The operational name of the venue
- Venue Postal Code: The postal code of the venue

The new dataframe will be created using both of the datasets above with the distance between the residence and medical venue, time travel to reach venue, 15minFlag which indicates whether, the distance can be travelled under 15 minutes.

3. METHODOLOGY:

Step 1: Import the Data

As described above, Foursquare for medical service venues, the open data DC project for social residences.

Step 2: Data Preparation

Here we will limit our project to research only the Downtown, Washington DC. This will allow us to control the pull request from Foursquare API because we will be using the basic foursquare API, where the request have a limit. And to manage the server memory. 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).

Step 3: Calculate minimal walking distances

In this step we create a new dataframe that has calculated columns where we calculate the distance between the every residence and the medical service and the minimum time travel and the flag column to indicate if the residence meets the 15 minutes criteria.

Step 4: Exploratory data analysis

Here we will do exploratory data analysis such as

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

Statistical analysis is done such as

- 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

Map Visualisation plots such as

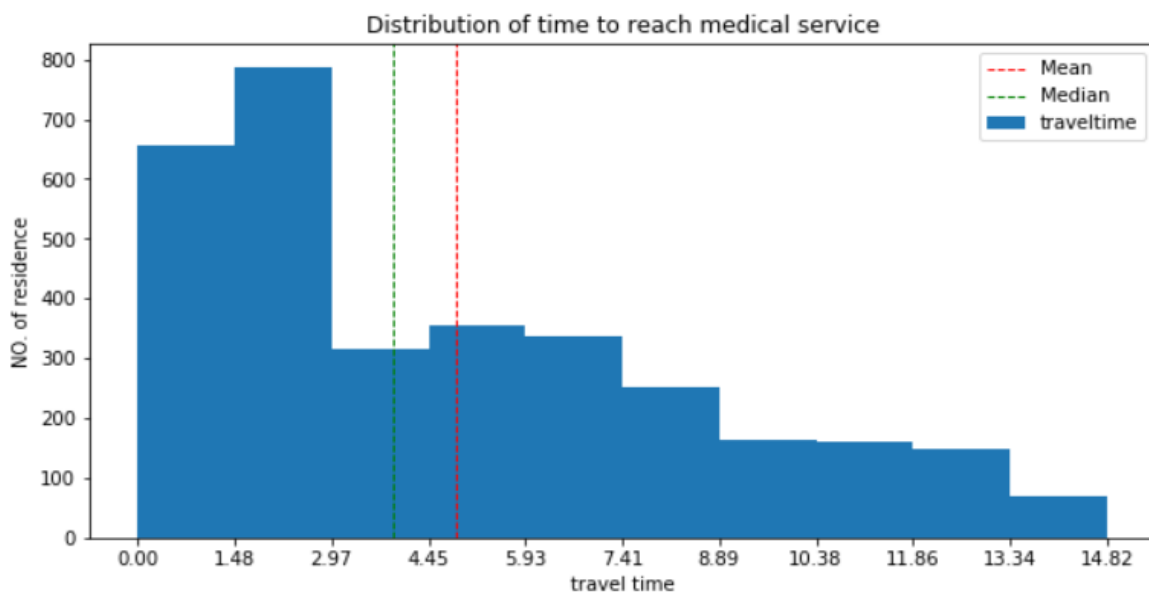
- Identify residences that meet or do not meet the 15 minutes criteria overall
- Identify number of distinct medical services within 15 minutes walk for each residence

Step 5: Unsupervised Clustering

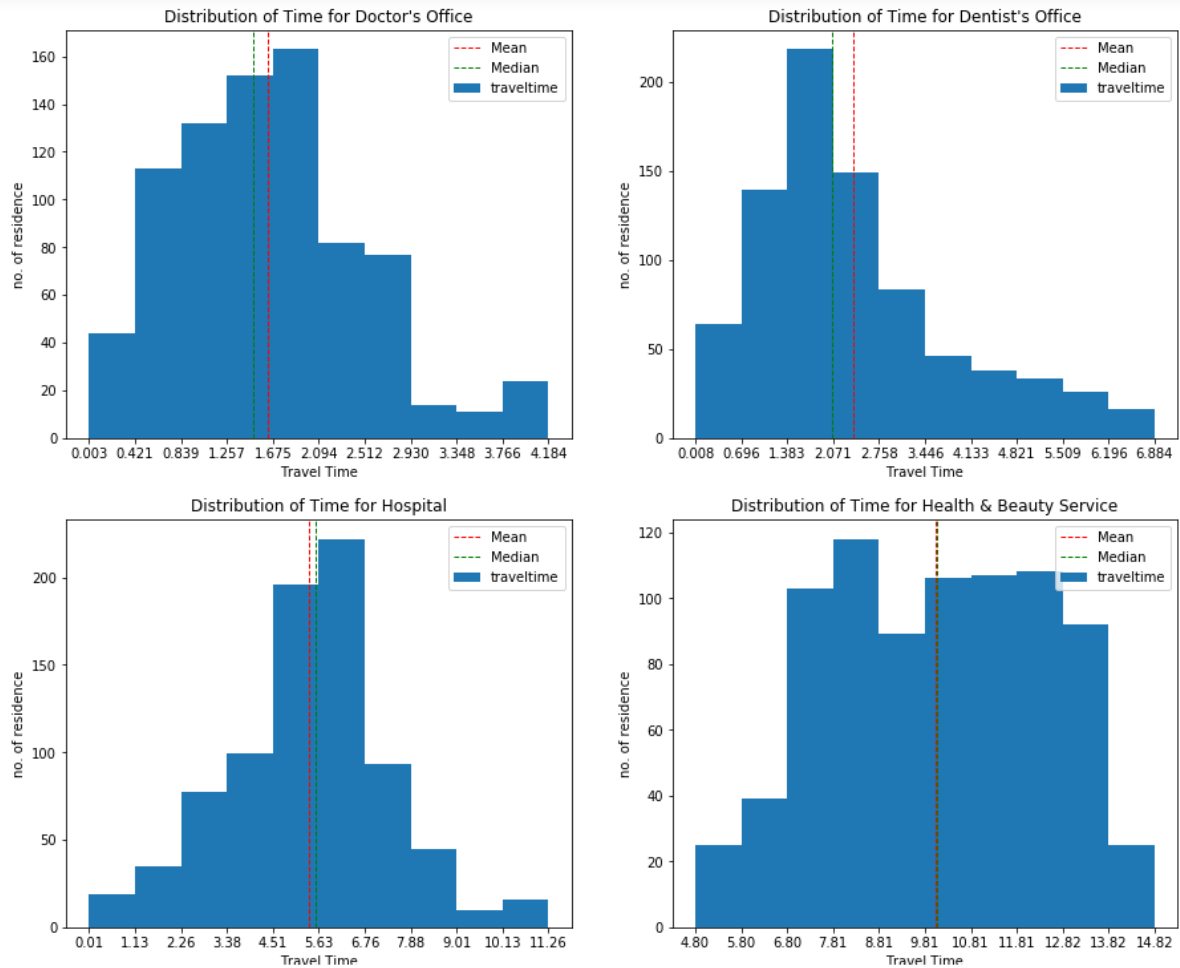
The reason to use clustering method is to group the residence that have common properties in term of 15-minute criteria. Interpreting the clusters, by confirming that they have a somewhat identifiable logic explaining why some residences belong to the same cluster, could help determine a specific approach 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. RESULT:

The exploratory data analysis shows that every residence present in our downtown dataset is within the range of 15 minutes of travel time by walking to reach to nearest medical service. To travel to the nearest medical service in downtown the average time is under 5 minutes.



To see the distribution of travel time for different medical service such as Doctor's office, Dentist's office, Hospital, Health & Beauty Service and the average travel time to reach the nearest venue in each medical service.

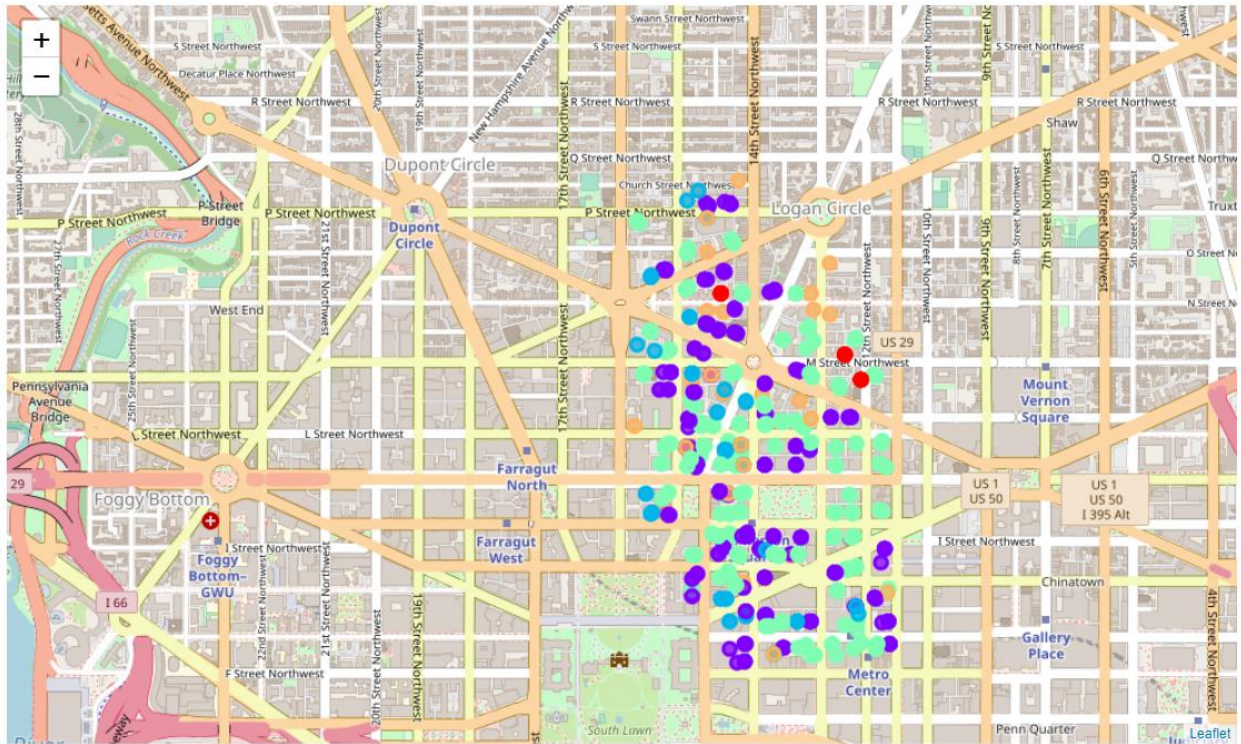


Here we observe that the distribution of the travel time to reach a Doctor's office is where we have maximum travel time of reach a doctor's office is under 5 minutes while the average time is 2 minutes.

For a more serious emergency where a person has to reach to a hospital, the maximum travel time taken to reach a hospital in downtown is 11 minutes and the average time is little over 5 minutes.

Note: The travel time calculated here is based on the linear distance instead of Manhattan distance, which is not the best measure of distance between two locations. The travel time does not take account of the traffic on road in the downtown. Traffic may vary depending on the time of day and the day of week.

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.



5. CONCLUSION:

- For the neighborhood that we selected for this project that is downtown, DC, we have been able to determine the "15 minutes" readiness of these neighborhood, 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.
- 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.
- We confirmed that every residence in the downtown, is within the reach of every type of medical service/category required within the 15 minutes of walk.

6. FUTURE WORK:

Assuming there is no limit for the foursquare API, we can expand the radius of search and include more neighborhood to the analysis with little changes to the code. 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. 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.