THE “15 MINUTES CITY” PROJECT

1. INTRODUCTION

1.2 BACKGROUND

In March 2020 mayoral elections will take place in Paris, France. One of the leading candidates wants to make Paris a city where anyone living in the city can reach key services and amenities within a 15 minutes 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 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 75018 is for the 18th district.*

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 two datasets generated from Foursquare and create a third dataset derived from the first two. First, let's review a few assumptions and decisions, before reviewing our data choices in detail.

2.1 IMPORTANT! 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. 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.

2. We will use a typical biking speed of 16km/h (~ 10mph). This speed would of course vary depending on the person's health and fitness, on the terrain and on traffic patterns.

   This conservative speed was selected after consulting this article about bike commuters:
   http://bikecommuterhero.com/whats-the-average-cycling-speed-of-a-bike-commuter/

3. 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:
   - Services to be included: a few categories of medical services (described in section 2.2 below)
   - Districts: 17th, 18th and 19th districts, located in the northern part of Paris

   However our data design approach and project methodology should be generic enough to include more services or districts.

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 **services_df**

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

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 services_df:

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 (17th, 18th and 19th districts as explained above) so that we can calculate their minimal distance to amenities included in our analysis.

Data source	Foursquare venues, limited to some sub-categories of the "Residence" venue category
Target data storage	A panda dataframe called **residences_df**

Venue categories and associated category IDs to be included:

Home (private)	4bf58dd8d48988d103941735
Residential building (apartment/condo)	4d954b06a243a5684965b473

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:

venues only
venue categories in the list of venue category IDs listed above
country = France
postal code = 75017 or 75018 or 75019 to only include addresses in the 17th, 18th and 19th districts of Paris

We will also calculate some new features in this dataset:

- District number (as explained above, by looking at the last 2 digits of the postal code).
- 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 = Boolean multiplication of 15MinFlag values	CALCULATED FEATURE 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 **res_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 res_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 services_df dataframe	
MinTravelTime	= Minimum of all travel times from a given ResID to all Services for a given ServCat	CALCULATED FEATURE 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 = 0 otherwise	CALCULATED FEATURE