

STUDY OF THE SURROUNDINGS OF THE METRO LINE 1 IN THE CITY OF PARIS, FRANCE

1. Introduction

1.1. Business problem

La Défense business district is a very attractive area located West of Paris. Many urban people work there and take the metro line 1 direct to this area every day. They likely ask where they could live in Paris with easy access to their work site that means they exclude losing time with transport connections. What places combine both a neighborhood with their favorite venues and a reduced transport time to go to work?

1.2. Objective

In order to answer that question, we will study the surroundings of all the stations of the metro line 1 for a business man who is looking for a home close to a metro entrance and close to restaurants and cinemas (for example).

So the project aims to cluster the metro 1 stations according to:

- (1) transport time between La Défense station and the home's station,
- (2) the numbers of restaurants and cinemas around the home's station.

2. Data presentation: acquisition and preprocessing

2. 1. Data source

- Metro line 1 stations

We can collect the coordinates of the metro stations on the RATP website: https://dataratp2.opendatasoft.com/explore/dataset/positions-geographiques-des-stations-du-reseau-ratp/export/?disjunctive.stop_name&location=9,48.86463,2.39738 (Licence ODbL Version Française)

We can determine the locations of all the stations and the distance between La Défense district and the potential home. So we have an estimation of the transport time.

- Venues

We can use the Foursquare application that provides us a list of the venues around each station in a radius of 300 meters. The result is not exhaustive because it depends

on the focused place. However we consider that the Foursquare database is enough supplied for a city like Paris.

2. 2. Data cleaning

- Stations

The downloaded csv file from RATP website includes a complete dataset about public transport of Paris. We select the data that interest us: the geographical coordinates and the names of the 25 stations belonging to the metro line 1. We drop all the other columns and rows, make the needed conversions of the data's types, and split the column « coordinates » into two columns « Latitude » and « Longitude ». We control the possible redundant or missing data.

We add a column « Transport time » based on the relative distance between each station and La Défense station. The metro line 1 has a W-E orientation, and La Défense station is the western end of the line. So we sort the stations by their latitude and we assign to each station a transport time that is proportional to its position on the line.

- Venues

For each station of the previous stations table, we make a request to the Foursquare app and get back all the nearby restaurants and cinemas (i.e. located in a radius of 300 m). We count the results for the two categories and group the total numbers in a global table.

Our dataset is now ready and contains 25 rows and 6 columns (cf. Fig. 1). The three variables are : Transport time, Number of restaurants and Number of cinemas.

Station	Latitude	Longitude	Transport time	Restaurant	Cinema
LA DEFENSE-GRANDE ARCHE	48.892187	2.237018	0	7	1
ESPLANADE DE LA DEFENSE	48.888631	2.247932	1	10	0
PONT DE NEUILLY (AVENUE DE MADRID)	48.884708	2.260515	2	1	0
LES SABLONS (JARDIN D'ACCLIMATATION)	48.881192	2.271687	3	1	0
PORTE MAILLOT	48.877551	2.283162	4	8	1

Figure 1. Extract of the stations dataset after cleaning: transport time and number of venues on metro line 1.