Data Science

Capstone Project

08.07.20

Introduction.

Case study

High-tech company
Set up a headquarter shift
In Paris from Petit-Montrouge neighborhood

Since work done from office
→ Mainly consider workers interest

Same characteristics neighborhood based on :
Venues in the area
Population attributes

Project description

Project : finding similarities in Paris neighborhood to inform the company decision to acquire new headquarters

Idea : clustering twice
On venues
On population attributes

Data required :
Paris neighborhoods list and geographical information
Paris neighborhoods most frequent venues
Paris neighborhoods population data

Data.

geographical information

As geographical_df

Source

open data paris - geographical information

Attributes

Id_seq (link to other data)
Code INSEE (idem)
Borough
Neighborhood
Latitude
Longitude

Venues data

As paris_venues

Source

Function getNearbyVenues from foursquare API
Foursquare - developers apps

Attributes

Neighborhood
1st Most Common Venue
...
mth Most Common Venue

Population attributes

As population_df

Source

open data aput - rencensement

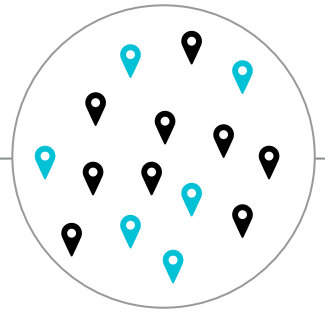
Attributes

Code INSEE commune
Density
Population
Population under 40

Idea diagram.

First clustering

Discriminate neighborhoods based on most common venues



Second clustering

Discriminate neighborhoods based on population information

Final decision

Choose neighborhood based on current possibilities (and again worker preferences if possible