

Coursera IBM Data Science Specialization Capstone Project

Report on the Study of Relocation to the Greater Paris Area

May 2019

Contents

1. Introduction	1
1.1 Business Problem and the Purpose of this Study	1
2. Data Acquisition	2
2.1 Stage 1 – Acquisition of postal codes, boroughs, neighborhoods and geographic coordinates	2
2.2 Stage 2 – Acquisition of venues with FourSquare	3
3 Data Analysis Methodology	3
3.1 Data preparation for clustering	3
3.2 Data Clustering	4
4 Results	5
4.1 Cluster meaning	5
4.2 Cluster Comparison	5
5 Conclusion	7

1. Introduction

Accommodation in the center of many big cities is in high demand due to the proximity to business areas, sightseeing, parks and other sites which are very convenient in the day-to-day life. Increasing demand for accommodation in city centers is very often accompanied with a rather limited offer and thus with an increase of the rent.

Many people can no longer afford living in city centers and have to relocate to suburbs. When doing so, they try to find a location which as much as possible corresponds to the place in the city center they used to live in.

1.1 Business Problem and the Purpose of this Study

This trend can be observed to a great extent in the **capital of France, Paris**. Paris is a rather small city as far as its area is concerned (around 15 kilometers – 10 miles in diameter). Accommodation has become more and more scarce inside Paris, and the rent has

considerably increased. More and more people are seeking to relocate to the Greater Paris Area where rent is not so high, and the offer is more diverse.

The purpose of this study is to assist people relocating to the Greater Paris area, provide them information where they can find approximately the same venues as in the city center.

2. Data Acquisition

Data acquisition and analysis were performed using Python programming language and its Pandas, Numpy, Scikit-learn and BeautifulSoup libraries among others.

Data was collected in two stages:

2.1 Stage 1 – Acquisition of postal codes, boroughs, neighborhoods and geographic coordinates

First, I acquired data for Paris

- Postal codes, boroughs and neighborhoods from https://fr.geneawiki.com/index.php/Liste_des_quartiers_de_Paris.

Then, I acquired data for the Greater Paris Area (Ile-de-France region)

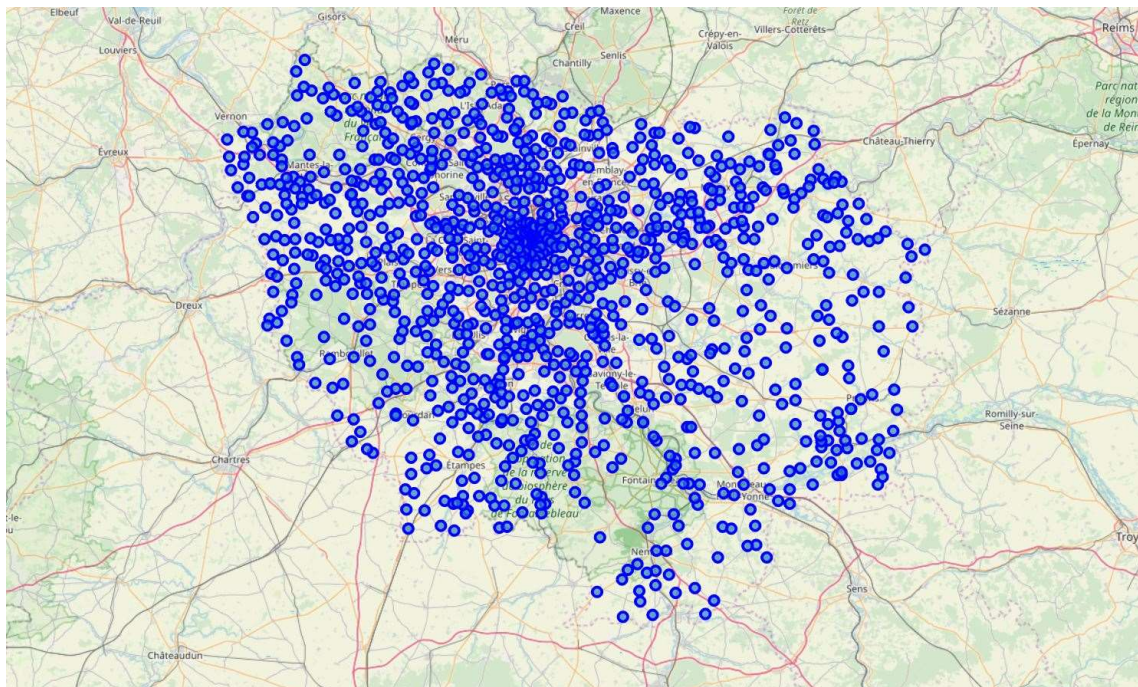
- Postal codes, boroughs and neighborhoods from <https://www.annuaire-administration.com/code-postal/region/ile-de-france.html>.

A Python library for parsing html – BeautifulSoup – was used for acquiring data.

- Geographic coordinates of neighborhoods were loaded with the geocoder Nominatim

Data for Paris and the Greater Paris area was merged for further analysis.

Map of Neighborhoods in Greater Paris Area



2.2 Stage 2 – Acquisition of venues with FourSquare

The FourSquare platform was used to load venues within a certain radius of each neighborhood as well as their geographic coordinates.

3 Data Analysis Methodology

3.1 Data preparation for clustering

Data was prepared using the following methods:

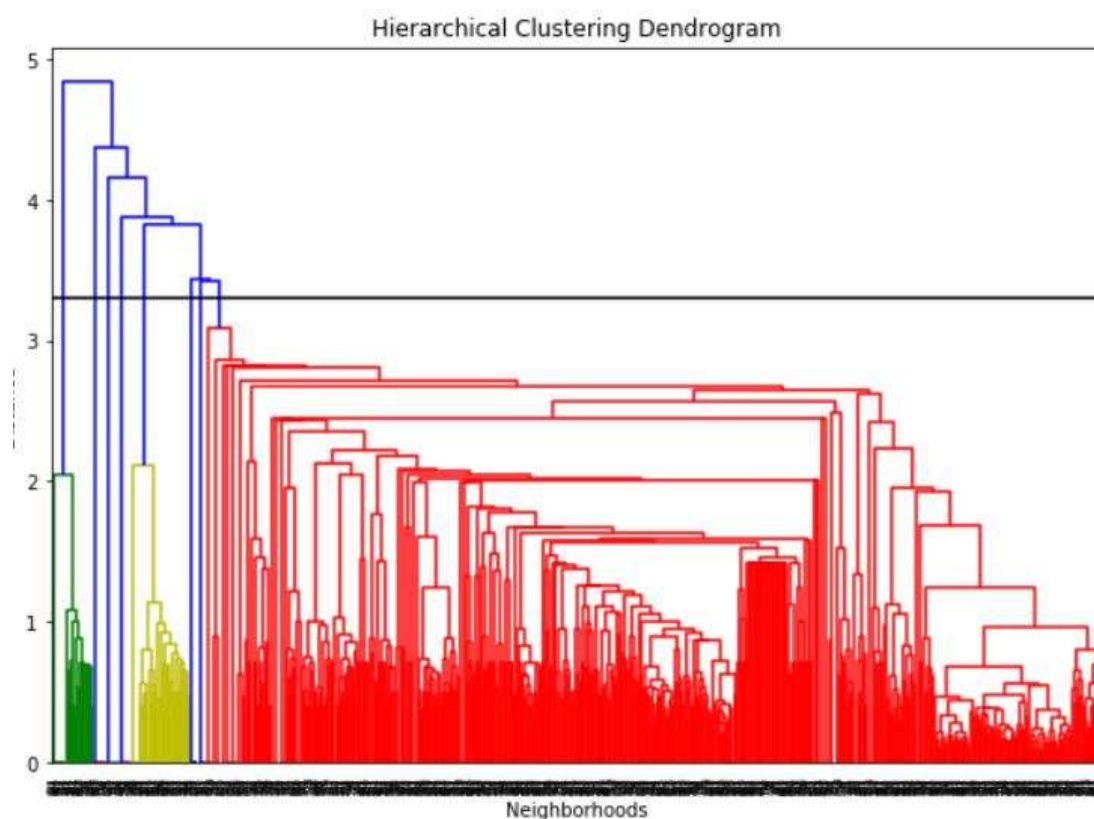
- One-hot encoding of venue types in each neighborhood
- Grouping rows by neighborhood and taking the mean of the frequency of occurrence of each category of venues
- Putting most common venues in a Pandas data frame

Top rows of the resulting data frame with venues sorted by their mean frequency of occurrence

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ablis	Photography Studio	Zoo Exhibit	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service	Event Space	Exhibit	Fabric Shop
1	Ablon-sur-Seine	Bank	Train Station	Hotel	Pharmacy	Zoo Exhibit	Falafel Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service
2	Achères	Sports Club	Locksmith	Italian Restaurant	Plaza	Beach	Farm	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service
3	Aigremont	Sporting Goods Shop	Golf Course	Pharmacy	Roof Deck	Falafel Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Event Service
4	Alfortville	Park	Restaurant	Music Venue	Flea Market	Supermarket	Middle Eastern Restaurant	Bus Stop	Outdoor Sculpture	Bakery	Caribbean Restaurant

3.2 Data Clustering

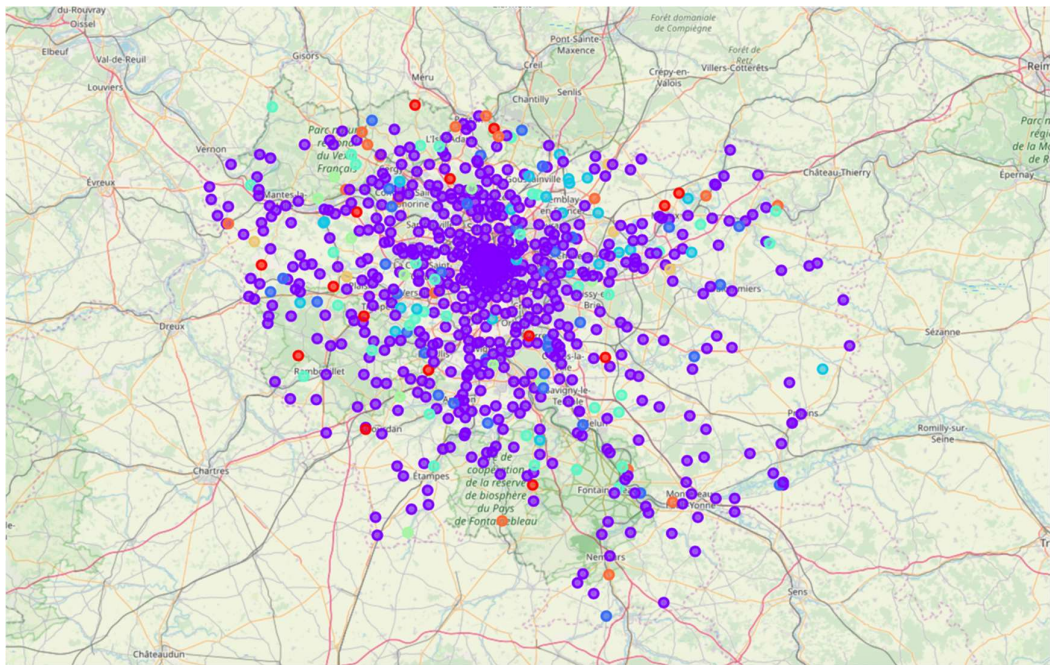
I used **K-Means** algorithm in order to cluster data. To determine the optimal number of clusters I used a dendrogram



If we draw a horizontal line at approximately 3.5, we get 8 clusters

I decided to cluster venues clustered into 8 clusters based on the above dendrogram. There is no fixed rule where to draw a line in the dendrogram, but 8 seems to be a logical choice since a bigger number of clusters might be difficult to interpret.

Types of venues in the Greater Paris Area



4 Results

4.1 Cluster meaning

Based on the most frequent venue types in clusters, we can extract the following meaning of clusters (give them titles).

Number	Title
0	Restaurants of any type of cuisine
1	French Restaurants and other catering facilities
2	Supermarkets
3	Hotels
4	French Restaurants (related to cluster 1)
5	Bakeries
6	Auto Workshops
7	Train Stations (public transport)

4.2 Cluster Comparison

A person relocating from Central Paris to the Greater Paris area (suburbs) might want to find approximately the same venues in the suburbs as he or she used to have in Paris. Our clusters contain neighborhoods in Paris and in the Greater Paris area. In order to guide this person where to go, I extracted the Greater Paris neighborhoods from each cluster. This person can find among them a neighborhood corresponding to his or her preferences based on the venues in the neighborhood this person is relocating from.

Selected neighborhoods per cluster type in the suburbs (for full details please refer to Jupyter notebook)

Number	Title	Neighborhood
0	Restaurants of any type of cuisine	Crégy-lès-Meaux Coubert Varreddes, Poigny-la-Forêt Chapet La Verrière...
1	French restaurants and other catering facilities	Auteuil Collégien Meaux Soignolles-en-Brie Coulommiers...
2	Supermarkets	Nanteuil-lès-Meaux Mouroux Lésigny Le Pin Sept-Sorts...
3	Hotels	Chalifert Villeneuve-sous-Dammartin Dampmart Saint-Thibault-des-Vignes Gressy...
4	French Restaurants (related to cluster 1)	Pontcarré Émerainville Avon Crisenoy Montceaux-lès-Meaux...
5	Bakeries	Croissy-Beaubourg Gargenville Les Alluets-le-Roi Bailly Saint-Germain-lès-Corbeil...
6	Auto Workshops	Dammartin-en-Goële Annet-sur-Marne Voulangis Dammartin-en-Serve Saint-Germain-de-la-Grange...

7	Train Stations (public transport)	La Grande-Paroisse Bagneaux-sur-Loing Compans Isles-les-Meldeuses Nanteuil-sur-Marne...
---	-----------------------------------	---

5 Conclusion

This study allowed me to compare neighborhoods in Central Paris with Neighborhoods in the Greater Paris Area using venue types for comparison. The results can be used as a guide by persons relocating from the Central Paris Area to the suburbs in order to find living conditions similar to those they used to have in the Central Paris. Taking into account that many people are currently relocating from central city areas to the suburbs and this trend is becoming wide-spread this study can be extended to other big international cities and the areas around them.