Capstone Project
Battle of the Neighborhoods

● ● ●

Chris Weber, 29.05.2020

Contents

1. Introduction - the business problem to be solved
2. Data section - description of data requirements and sources
3. Methodology section:
   a. Main component of the report
   b. Data processing
4. Results - description of the results
5. Discussion - observations and further research suggestions

1. Introduction

Background

Paris is one of the fashion capitals of this world. Many fashion retailers that want to open a boutique in Paris will want to know where the best location could be for their specific fashion business and category. With best location we mean the area in which to expect the highest income from retail sales for a given product.

Problem to be solved

The goal of this analysis will is to find the frequency per each retail category, i.e. shoes, women's clothing etc and cluster neighborhoods by category of fashion retailers. Therefore retailers will be able to understand, which area should be best for opening a store with a given clothing type, e.g. Women's Clothing, Shoes etc.

2. Data Section

Data requirements:

-   A list of all venues in Paris. For this we will use the Foursquare Places API through which we will get names, geo locations and retail categories for all venues in Paris. However we will filter this list to only show venues that fall into the category "Clothing Store".
-   A list of neighborhoods in Paris. For this list we will scrape the website "https://en.wikipedia.org/wiki/Arrondissements_of_Paris". From this page we will read out names, population density and any other information about the arrondissements of Paris that might be helpful for our analysis
-   A list of the geo locations of the Paris neighborhoods. For this we will translate Paris zip codes for each neighborhood into geo locations using the Nominatim Python package

2. Data Section

Data processing & tools used:

- Paris neighborhoods are obtained using Wikipedia. Zip codes for each neighborhood correspond to their numbering, e.g. 1st arrondissement has zip code 75001, 2nd has 75002 and so on. Therefore we can fill in these details manually.
- Paris map is created using Nominatim and Folium Python package
- Folium will be our main tool visualize all acquired data
- K-Means Clustering will be used to cluster our venues

3. Methodology

Approach

The goal is to map the data described in section 2.0 in order to facilitate the choice of neighborhood for opening a clothing store. The information will be consolidated in one map where you can see the location of neighborhoods and the cluster of clothing stores in the neighborhood.

Tools

Web-scraping of Wikipedia was used to create a data-frame creating all information we need about the Paris neighborhoods. Geodata was obtained using Nominatim Python package to get latitude and longitude of all neighborhoods in Paris. Folium was used to create a map of Paris, show markers for each neighborhood and color these markers based on the cluster labels.
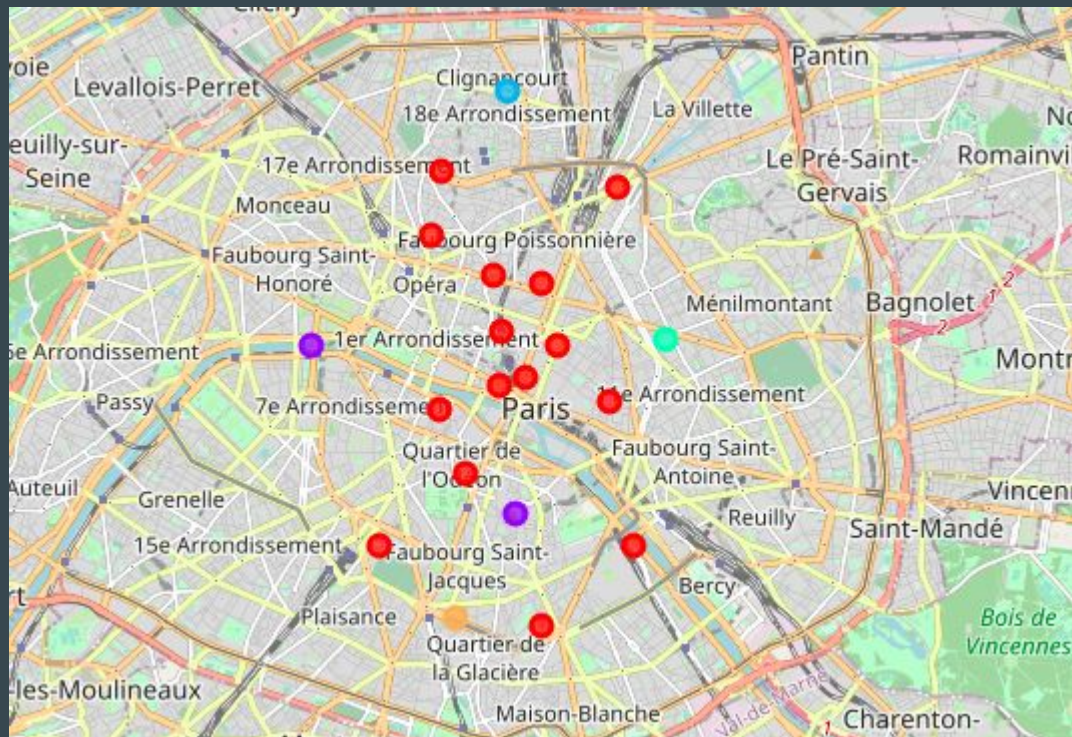
4. Results

Neighborhoods with highest frequency for a given retail category:

Retail	Women's Store	Shoe Store	Men's Store	Kids Store	Clothing Store	Lingerie Store	Boutique
Area	Observatoire	Panthéon / Passy	Panthéon / Reuilly	Butte-Montmartre / Passy	Ménilmontant	Vaugirard	Opéra
Freq	0.67	0.50	0.50	0.50	0.83	0.22	0.22

4. Results

Paris map - Neighborhoods and clusters of venues:

4. Results

Cluster 2 (violet) - this cluster is great for opening a shoe store:

```
In [130]: paris_merged.loc[paris_merged['Cluster Labels'] == 1, paris_merged.columns[[1] + list(range(5, paris_merged.shape[1]))]]
Out[130]:
```

	Area (km2)	postal_code	latitude	longitude	Cluster Labels	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	
4	2.541 km2 (0.981 sq mi)	75005	48.8448	2.34795	1	Shoe Store	Men's Store	Women's Store	Jewelry Store	Adult Boutique	Arts & Crafts Store	Baby Store	Boutique	
15	16.305 km2 (6.295 sq mi)³7.846 km2 (3.029 sq mi)4	75016	48.8637	2.31285	1	Kids Store	Shoe Store	Women's Store	Adult Boutique	Arts & Crafts Store	Baby Store	Boutique	Clothing Store	De

4. Results

Cluster 3 (turquoise) - this cluster is great for opening a clothing store for kids:

```
In [131]: paris_merged.loc[paris_merged['Cluster Labels'] == 2, paris_merged.columns[[1] + list(range(5, paris_merged.shape[1]))]]
Out[131]:
```

	Area (km2)	postal_code	latitude	longitude	Cluster Labels	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
17	6.005 km2 (2.319 sq mi)	75018	48.8925	2.3466	2	Kids Store	Optical Shop	Women's Store	Adult Boutique	Arts & Crafts Store	Baby Store	Boutique	Clothing Store	Department Store

5. Discussion

While already being able to cluster neighborhoods based on the retail store each neighborhood contains, it would be interesting to add additional information like foot traffic and rental costs. This would enable retailers to decide not only based on frequency of clothing stores, but also make an assessment of how expensive rental costs would be compared to the expected foot traffic for each neighborhood. Therefore this would be the suggested next step for a deeper analysis.