Predicting the best Gilets Jaunes' demonstration localization to limit material damages in Paris

1. Business Plan

It has been now 5 months that *Gilets Jaunes* are demonstrating in Paris. Some are pacific whereas an increasing number of "black blocs" are decimating the old neighborhoods of the town, trashing streets and buildings and specifically targeting Gastronomic Restaurants, Luxury shops and other so-called symbolic places. The global cost for France is currently above the hundred million of euros.

In this project, we will try to identify the neighboorhods that are the most likely to be subject to vandalism and try to find neighborhoods where the demonstrators could be headed over in order to avoid damage as much as possible.

We believe that it can help the French Government and Paris citizens to estimate what damages could be done in case of a march. It is also essential for the *Gilets Jaunes* who are truly willing to speak their voice during a planned and government accepted-march, and want to demonstrate without the violence and degradations that had accompanied them every Saturday.

We make this project public and we know that there are plenty of other factors to take into account on this very sensitive subject. This project is just a preliminary to further analyses.

2. Data

We will use the borough and neighborhood data to find with Foursquare the venues in each neighborhoods. For that, we need a table with the boroough, neighborhood, coordinates and venues. First, we will have to download the JSON file of the Paris Neighborhoods at this address on **opendata.paris.fr** (link in the notebook). We will have to also take into account the kind of population living in these areas, who could join or not the movement, be more impacted, etc.

If the dataset were easily available, it would be nice to add to this datatable the index of criminality in the different neighborhood, as well as a population kind (student, residential, ...) index.

Instead, we will use Poverty Index and Median Life Level in euros known in Paris Borough (**INSEE 2015**) (link in the notebook).

The dataframe extracted is the following (5 first rows presented):

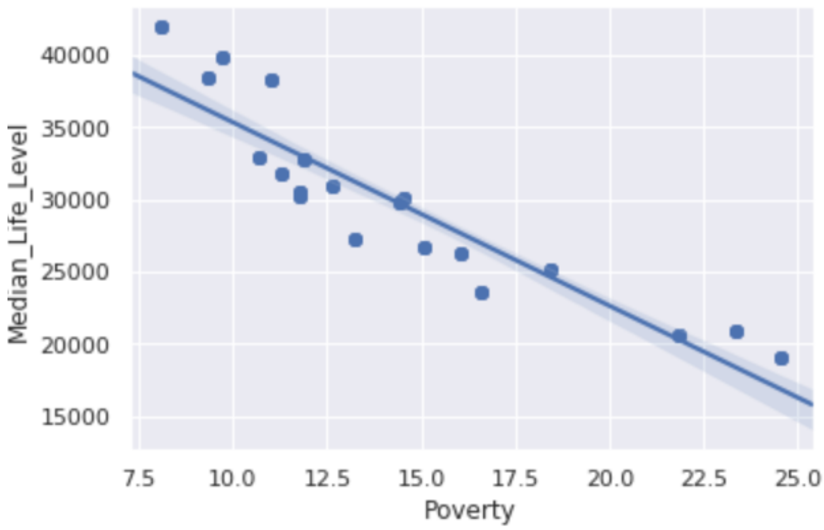
Borough	Neighborhood	Latitude	Longitude	Poverty	Median_Life_Level
3	Enfants-Rouges	48.863887	2.363123	12.643554	30988.000000
4	Notre-Dame	48.852896	2.352775	11.763913	30514.666667
5	Jardin-des-Plantes	48.841940	2.356894	10.678701	32950.000000
7	Saint-Thomas-d'Aquin	48.855263	2.325588	8.100699	41949.000000
9	Faubourg-Montmartre	48.873935	2.343253	11.894139	32771.000000

After that, we will find the coordinates associated with each neighborhood, and we will find the venues using Foursquare. Once we have the venues, we will try to cluster the neighborhoods in order to know which are the more likely to have huge damages.

3. Methodology

First, we clean the data to obtain the above dataframe. We will have to verify if Poverty and Median Life Level are correlated in order not to take the same parameter twice into account.

As we can notice, the relation between poverty and Median_Life Level is not linear and seems



logarithmic. Therefore, we decide to keep both of these indicators.

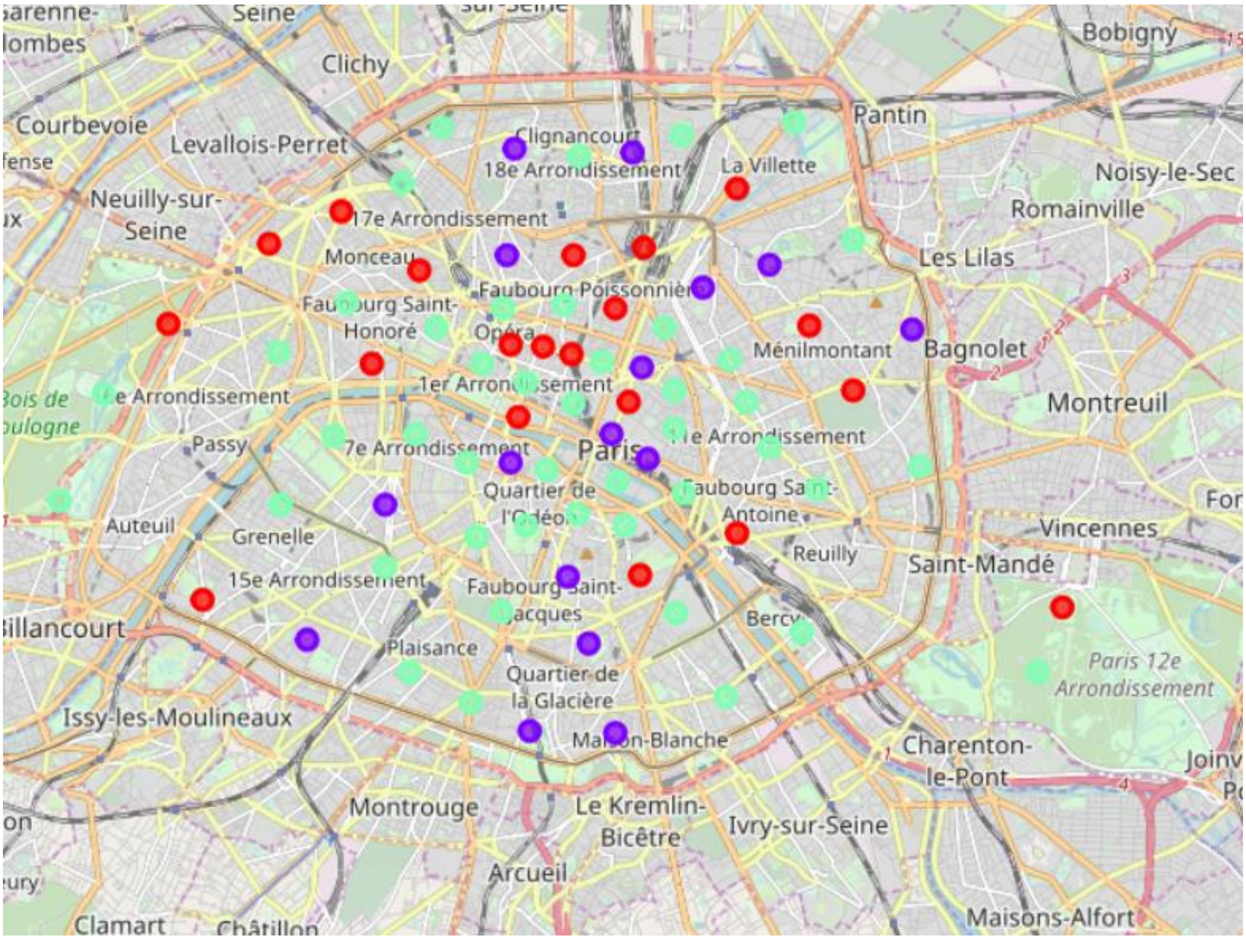
Using Foursquare, we will identify the most common venues in each neighborhood and along with poverty index and median life level, we will use K-mean clustering, which is interesting in a sense that we want to obtain an indicator of the propensity of a place to be damaged by

Black Blocks, therefore, we will have 3 clusters indicating Low-Moderate-High probability of been damaged. The more luxury shops and gastronomic restaurants, and poverty index is low, the more damage will be high. We consider that less rich places would be less willing to deteriorate their own area instead of a more symbolic one.

Results

K-mean clustering:

After creating 3 clusters based on the most common venues and the social indexes, we find these clusters, shown on the map :



Red:

Neighborhood	Median_Life_Level	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
Jardin-des-Plantes	32950.000000	0	French Restaurant	Hotel	Science Museum	Garden	Museum	Café	Botanical Garden	Bakery	Greek Restaurant
Rochechouart	32771.000000	0	French Restaurant	Hotel	Italian Restaurant	Bakery	Bistro	Coffee Shop	Vegetarian / Vegan Restaurant	Music Venue	Bar
Porte-Saint-Denis	25154.000000	0	French Restaurant	Bar	Pizza Place	Coffee Shop	Bistro	Italian Restaurant	Café	Turkish Restaurant	Burger Joint
St-Germain-l'Auxerrois	31842.555556	0	French Restaurant	Hotel	Café	Plaza	Historic Site	Exhibit	Cosmetics Shop	Art Museum	Bar
Ternes	29872.000000	0	French Restaurant	Hotel	Italian Restaurant	Bistro	Seafood Restaurant	Asian Restaurant	Pub	Restaurant	Moroccan Restaurant

Green:

Neighborhood	Median_Life_Level	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
Enfants-Rouges	30988.000000	2	Art Gallery	Wine Bar	Clothing Store	Café	Bistro	Hotel	Coffee Shop	Sandwich Place	Boutique
Notre-Dame	30514.666667	2	French Restaurant	Ice Cream Shop	Japanese Restaurant	Bakery	Plaza	Historic Site	Wine Bar	Creperie	Garden
Saint-Thomas-d'Aquin	41949.000000	2	French Restaurant	Hotel	Café	Italian Restaurant	Bakery	Coffee Shop	Historic Site	American Restaurant	Tailor Shop
Faubourg-Montmartre	32771.000000	2	French Restaurant	Hotel	Wine Bar	Thai Restaurant	Burger Joint	Italian Restaurant	Cocktail Bar	Creperie	Bar
Porte-Saint-Martin	25154.000000	2	French Restaurant	Coffee Shop	Hotel	Theater	Bar	Pizza Place	Italian Restaurant	Breakfast Spot	Café

Purple:

Neighborhood	Median_Life_Level	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
Combat	19136.956522	1	French Restaurant	Restaurant	Pool	Italian Restaurant	Bar	Park	Plaza	Bike Rental / Bike Share	Thai Restaurant
Ecole-Militaire	41949.000000	1	Hotel	French Restaurant	Plaza	Diner	Bistro	Farmers Market	Café	Garden	Dessert Shop
Saint-Georges	32771.000000	1	Hotel	French Restaurant	Cocktail Bar	Italian Restaurant	Bar	Theater	Café	Lounge	Japanese Restaurant
Maison-Blanche	23538.181818	1	Hotel	French Restaurant	Plaza	Bistro	Café	Diner	Garden	Supermarket	Bakery
Parc-de-Montsouris	27233.000000	1	Italian Restaurant	French Restaurant	Restaurant	Bus Stop	Middle Eastern Restaurant	Café	Sushi Restaurant	Park	Theater

With the data we have and the knowledge of the city, we could hypothesize that the “red” neighborhoods are the most exposed to damages since they have a higher frequency of French Restaurants and Hotels than the other clusters. But then, how to differentiate the 2 last clusters ?

This is unsatisfactory since there are a lot of incoherence. For example, why are PicPus and Bel-Air in different clusters? (see notebook). Thus, it seems that the clusters are somehow difficult to assess and might not be as accurate as wished.

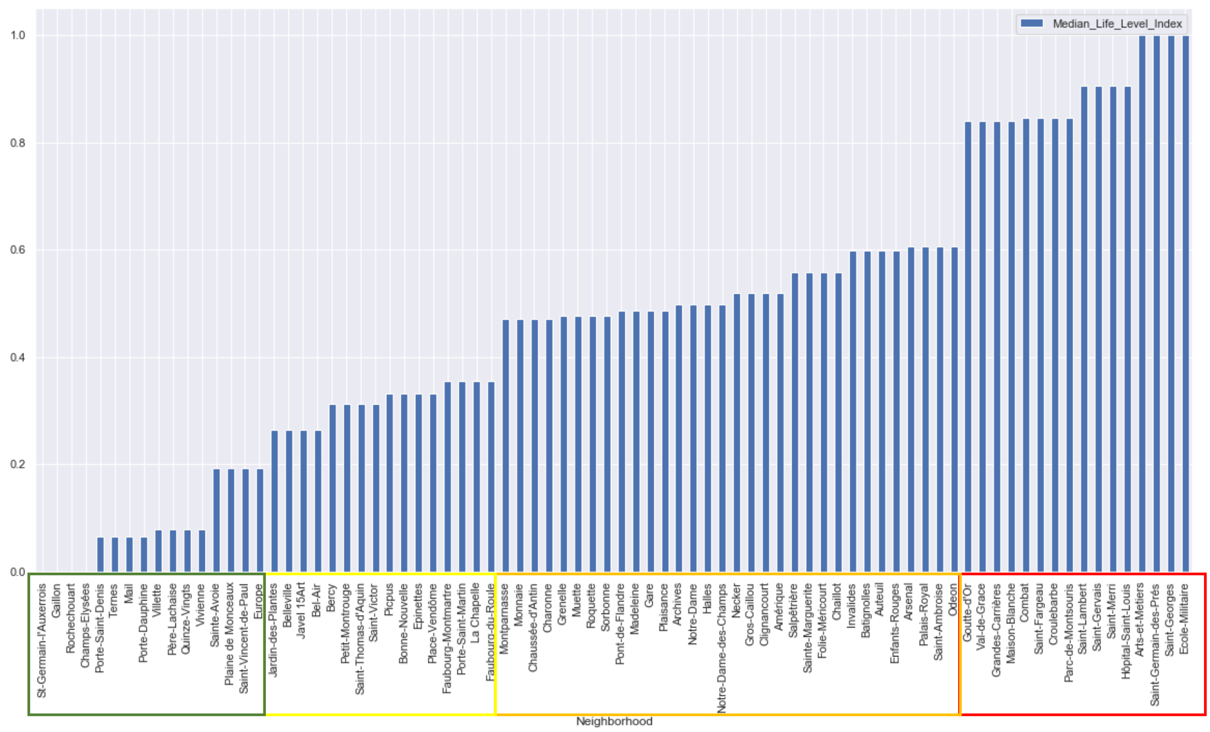
Neighborhoods contain similar venues and moreover, the Median Life Level Index is variable within a cluster.

Indeed, how make the difference between a common street French food restaurant and a gastronomic one, without more information? How to know it a shop is a luxury one or not? Is it a 1* hotel or a 4* one? Unfortunately, we don’t have the appropriate dataset to assess that.

Therefore, since the Median Life Level Index could prove of a certain wealth of neighborhoods, we will try to apply supervised algorithm using it as a label. We will figure this out in the following analysis:

First, we will encode the different levels of the Median Life Level Index, based on the following observation :

Median Life Level Index in Paris Neighborhoods (ascending order)



This encoding permits easier processing in the following algorithms : K-nearest neighbors, SVM, Decision Trees and Linear Regression.

Unfortunately, the accuracies are VERY low and we cannot rely on these predictions …

Algorithm	Jaccard	F1-Score
kNN	0.1875	0.128571
Decision Tree	0.3125	0.302083
SVM	0.1875	0.059211
Logistic Regression	0.1875	0.059211

Discussion & Conclusion

Knowing the venues in a neighbourhood is not sufficient to know the probability of huge damage during a demonstration, based on the hypotheses that damages are proportional to the shown wealth of streets, shops and restaurants. We need to have further information about their turnover and luxury index, as well as criminality data.

As a conclusion we having a better dataset would help arriving to a precise delimitation of neighborhoods at risk. In this study, we have only made the hypothesis that the cluster 0 might be the more at risk, BUT the accuracy of this possibility is very low.

Let's do further analysis when we will have time (unfortunately, I cannot finish it for the moment but I hope I could find a proper solution to this complex and burning issue)!