Methodology

To make this report and analyze the data I have, I started looking for the better information available about the city's areas. Which one of the the different divisions was more suitable for the analysis?
I started making a dataframe retrieved from Wikipedia, about the city's boroughs. But we I displayed them I the map, the distribution wasn't equal. I needed a better distributed and more frequent location area. So I thought of the Postal Code regions, but again, it had a very irregular distribution. So finally I found in the city's database a sheet of all the neighborhoods, with its outline written in polygon locations. So with shapely python module I got the centroid of each neighborhood. With that I got resolved the city location data, well distributed.

Then I used the data frame of each neighborhood to get from Foursquare the venues of each neighborhood in a certain distance(500 mts). With that information, I grouped with 'groupby' method and counted how many venues were in each neighborhood, and then how many of them were "Pizza Places" category, and sorted.

	Neighborhood
Venue Category	
Café	78
Argentinian Restaurant	74
Pizza Place	68
Coffee Shop	47
Ice Cream Shop	44

With that I could present a first element for the results of the report. But then I did the same search of venues, but with the search query "Pizza", to get all the pizza places only, per neighborhood.

Neighborhood	Venue Category
VILLA CRESPO	8
BELGRANO	6
SAN TELMO	6
SAN NICOLAS	6
VILLA URQUIZA	3

This let me use ‘groupby’ again to make a sorted descending list of the neighborhoods where there are more pizza places.

Then I continued with the dataframe of the general venues of the city, making a ‘one-hot’ encoding with ‘get_dumies’ method. To that dataframe I calculated the mean of each category in each neighborhood. This let me apply a ‘k-means clustering’ from ‘scikit-learn’ module. I clustered the city venues in five clusters. Then visualized it in a map of the city with folium module.