Methodology

I started looking for the better information available about the city's areas. Which one was more suitable for the analysis? I made a dataframe retrieved from Wikipedia, about the city's boroughs. I displayed them on a map and saw that the distribution wasn't equal. I needed a better distribution and more frequent location area.

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

Venue Category	Neighborhood
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 neighbourhood.

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 neighbourhoods 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.