



# THE BATTLE OF NEIGHBORHOODS

Applied Data Science Capstone Project  
COURSERA







# INTRODUCTION

# Introduction

- Bogotá, the capital of Colombia, is a city with more than 10 million habitants, it is the main economic center of the country. It is the Colombian city with the largest number of foreign companies, one of the factors that position it as the largest labor market in its country. Bogotá is the largest business platform in Colombia, with 21% of the companies registered in the country, and additionally, 67% of Colombian high-impact enterprises are in the city. Bogotá one of the five most attractive Latin American capitals to invest
- Despite being a very populated capital, it is one of the few in the world that for the number of habitants it has not a massive transport like the subway. Until this year 2020 begins the first works. That is why in the middle term economic growth is planned at all levels.

BUSINESS PROBLEM



# BUSINESS PROBLEM

- The problem is to identify the areas in which there are more businesses, which have been the most popular and which areas are in which there are more opportunities to open a business.

# DATA ACQUISITION

# Data acquisition and cleaning

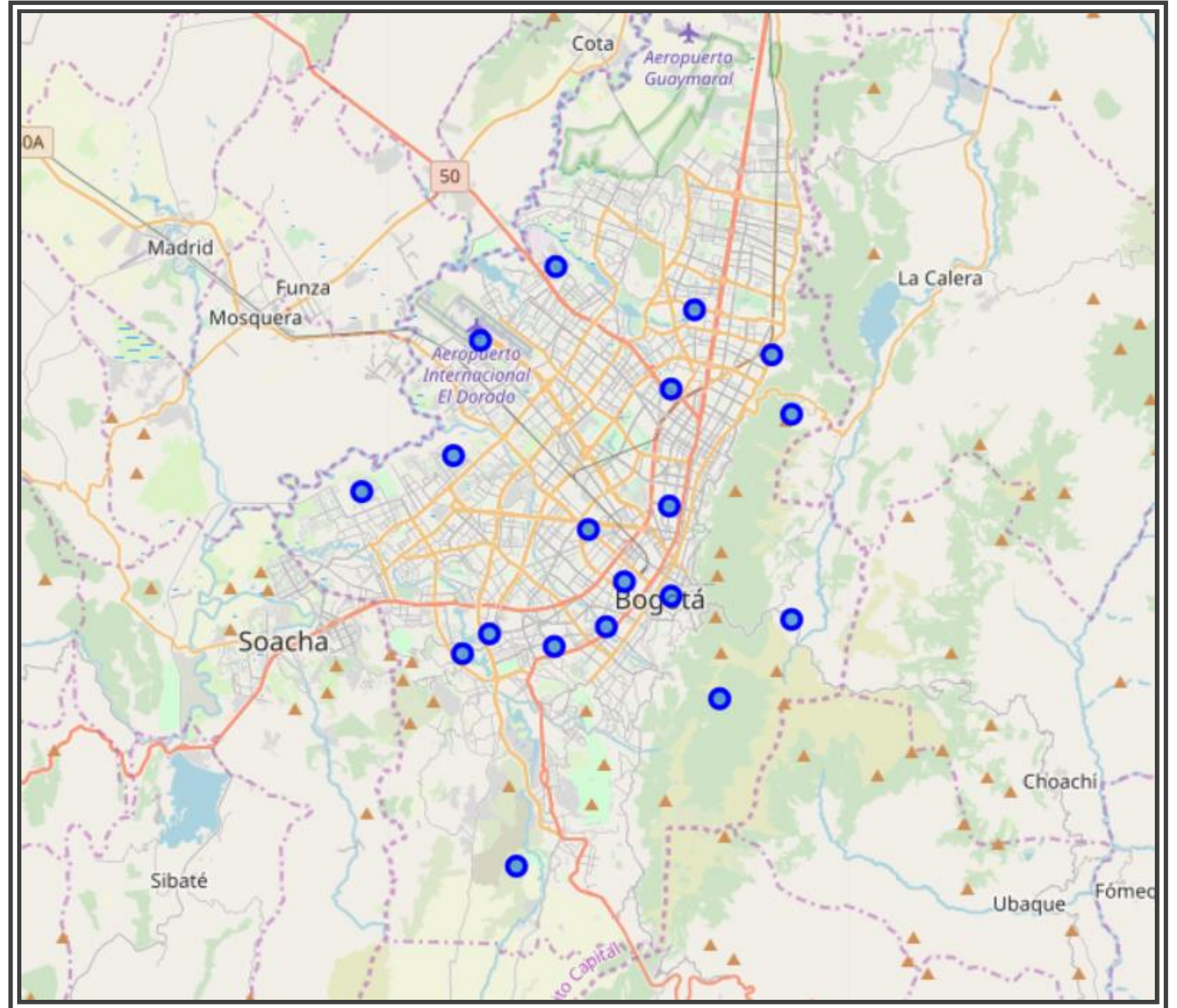
- To perform this analysis, the following main sources of data are available:
- Wikipedia, where we will get the locations in which the city is divided. Bogota is divided into 20 Neiborhoods or "Localidades". For each "Localidad" we will obtain the postal codes from this page: [https://es.wikipedia.org/wiki/Anex:Bogot%C3%A1\\_Localities](https://es.wikipedia.org/wiki/Anex:Bogot%C3%A1_Localities).
- Using the postal codes we can get the geographic coordinates. To do this is used geopy.
- With the previous data, we started the search of the different venues, the top venues categories in the different areas using the Foursquare API.
- ML tools will also be used as clustering to support the conclusions.
- Finally, we have the tools and data to start with the analysis.



METHODOLOGY

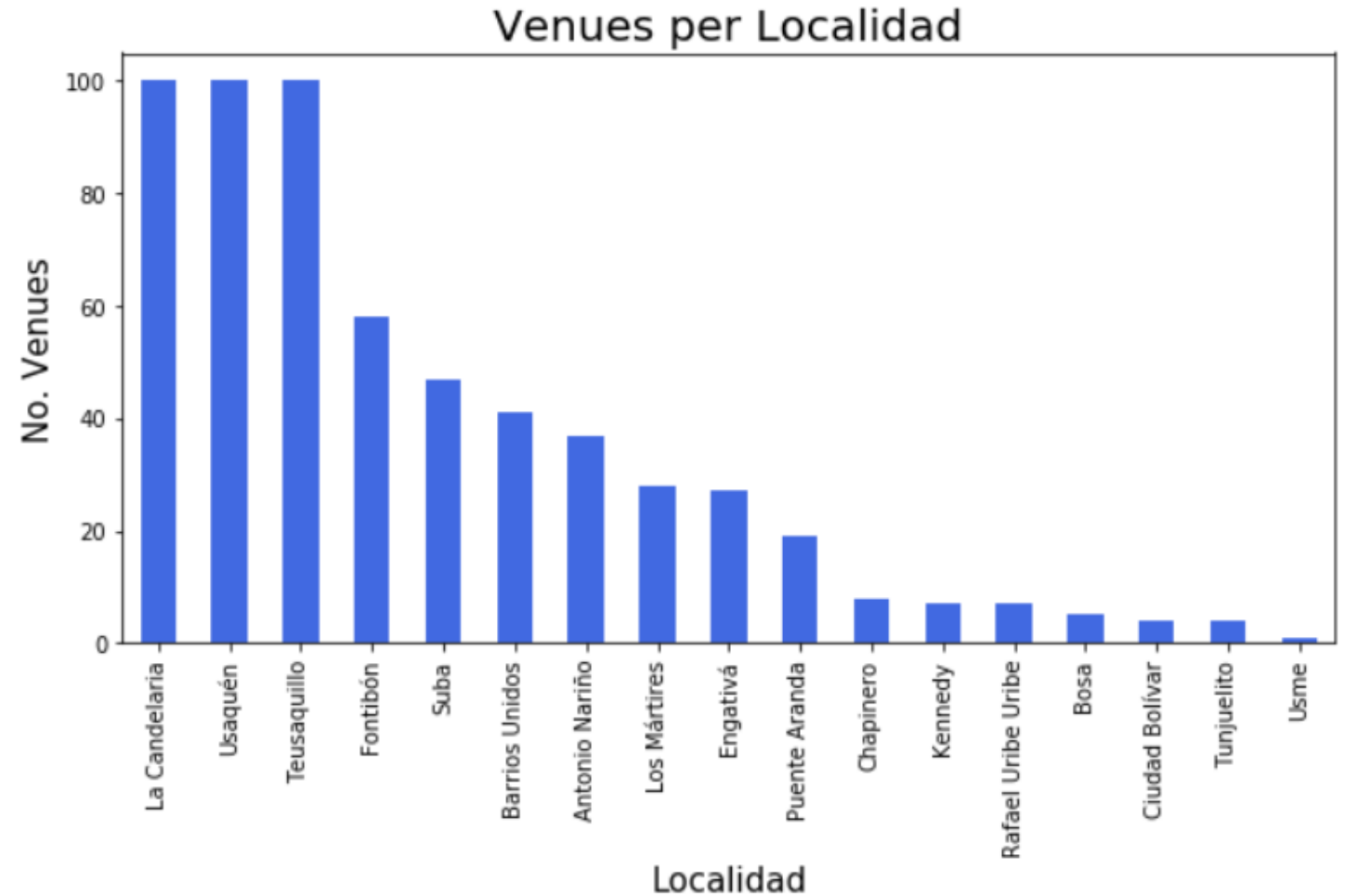
# Neighborhoods

- First, we need to do the geographical ubication of the centers for each neighbohood (Localidad) in Bogotá City

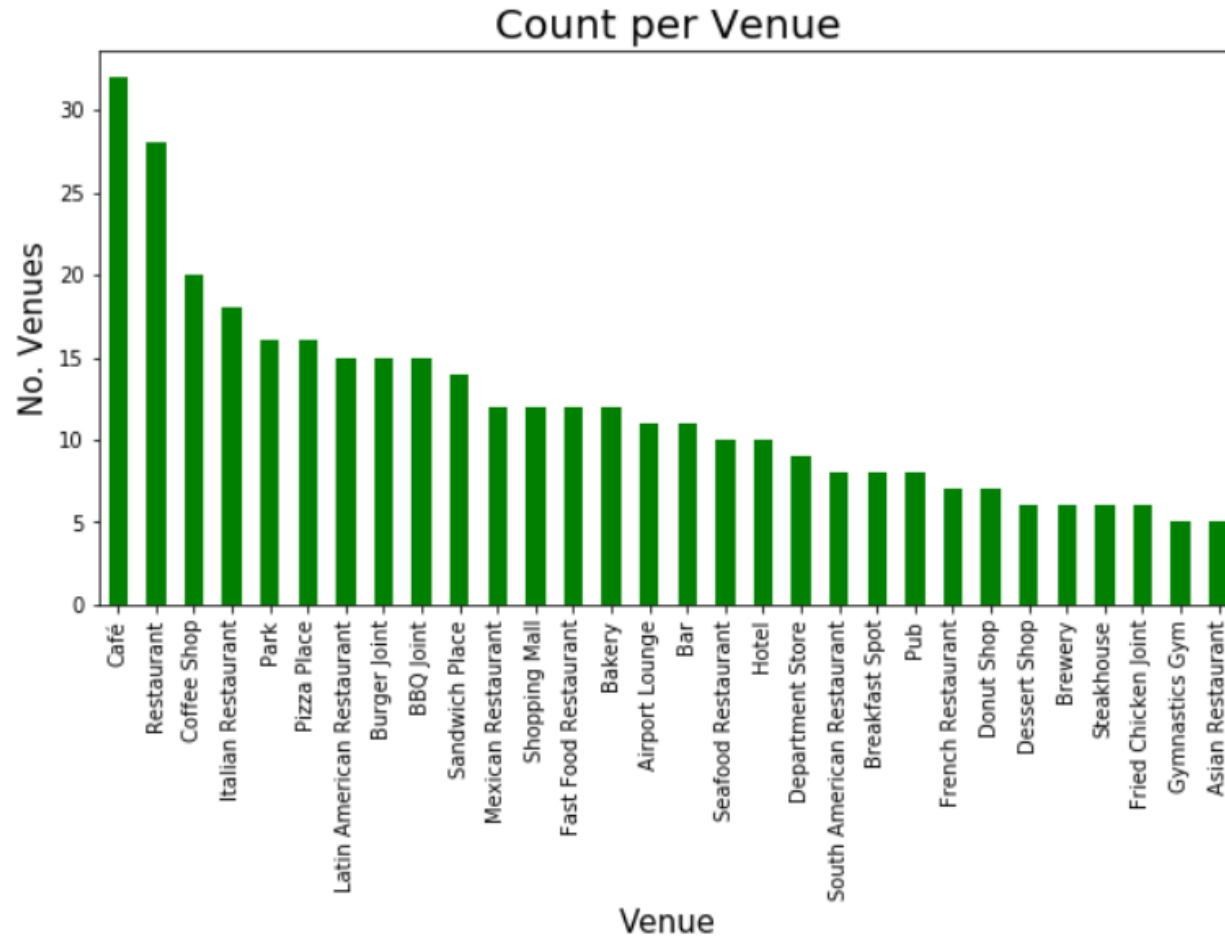


## Neighborhoods vs Venues

- Next, we identified the “Localidad” and the amount of venues in each one.
- We see that “La Candelaria”, “Usaquén” and “Teusaquillo” are the “Localidades” with highest amount of venues.

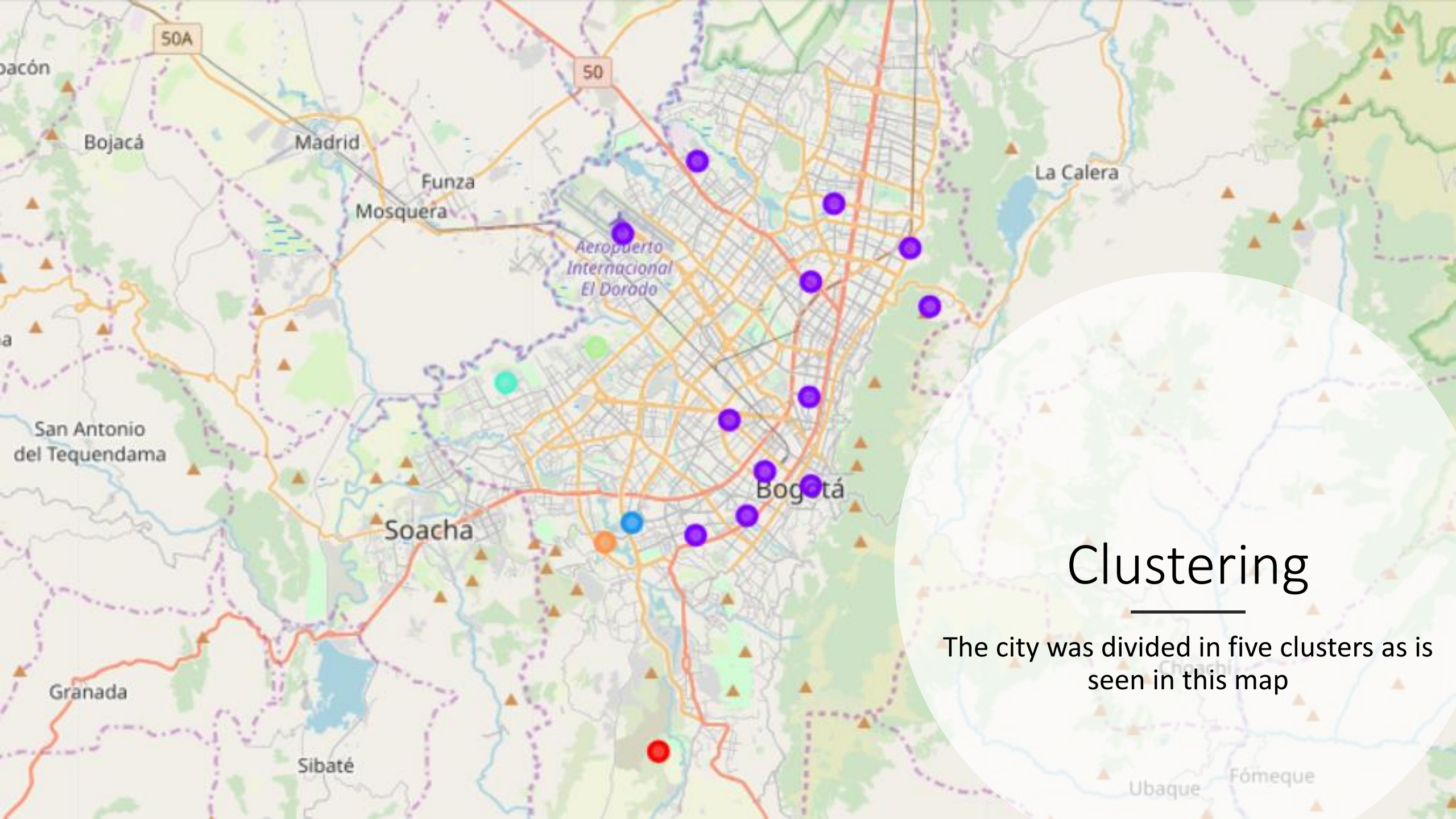


# Venues



- In this histogram we can see that coffes and Restaurants are the highest venues.
- It is striking that the parks are within the top 5 of Venues in Beogotá

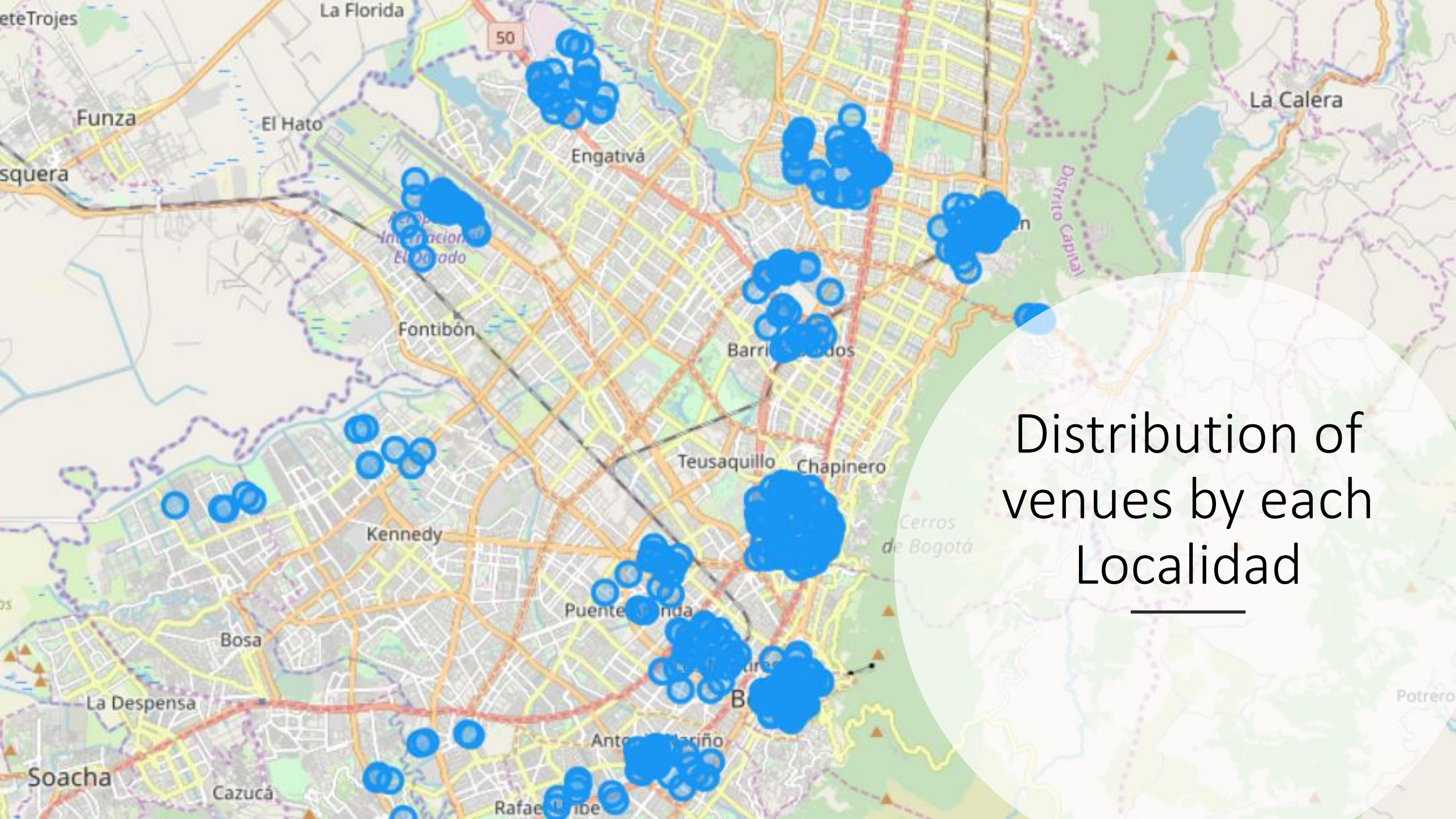




# Clustering

The city was divided in five clusters as is seen in this map

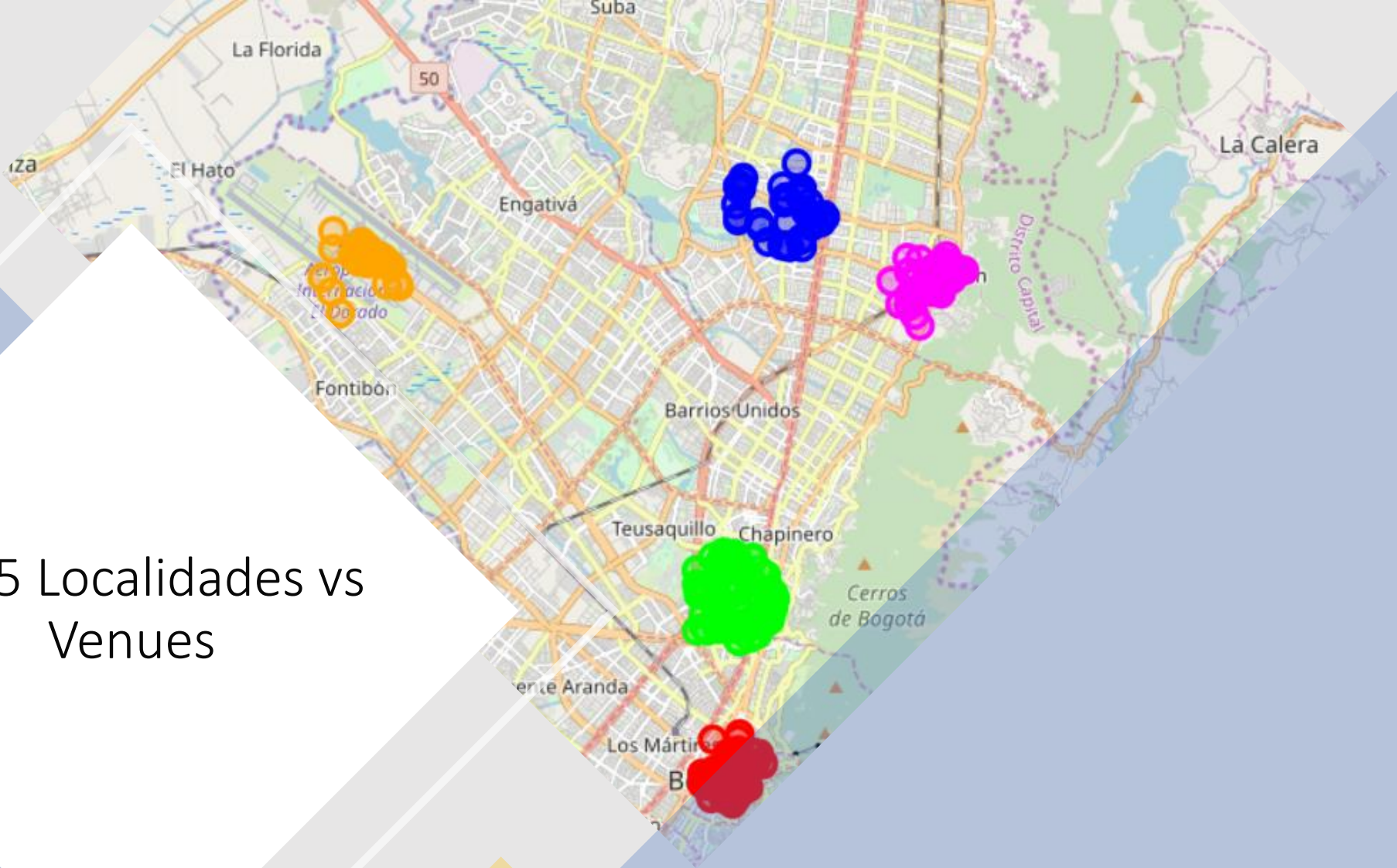




Distribution of  
venues by each  
Localidad



## Top 5 Localidades vs Venues



# Conclusion

- The Localidades with more venues are:
  - 'La Candelaria', 'Teusaquillo', 'Usaquén', 'Suba', 'Fontibón'
- It is interesting to note that these Localidades are not the most densely populated but in those places are main for offices and for tourist interest.
- Although there are many Venues, these sites still have the potential to open new sites such as gyms and multicultural restaurants
- We can recommend to analyze “La Candelaria” because is a in the town and it is a tourist site per excellence in which there is great potential for economic growth, and the future "Metro" will pass through this area.