

# BATTLE OF NEIGHBORHOODS

---

Where to open a new coffee shop in NYC

# Topic of the study

- A coffee shop owner in Flatbush, Brooklyn wants to open a another coffee shop in New York City
- He wants the decision on where to open this new shop to be data driven
- Study aims to determine neighborhood that present :
  - Similarity to Flatbush in terms of surrounding venues
  - Higher life standard than in Flatbush
  - A deficit of coffee shops compared to Flatbush

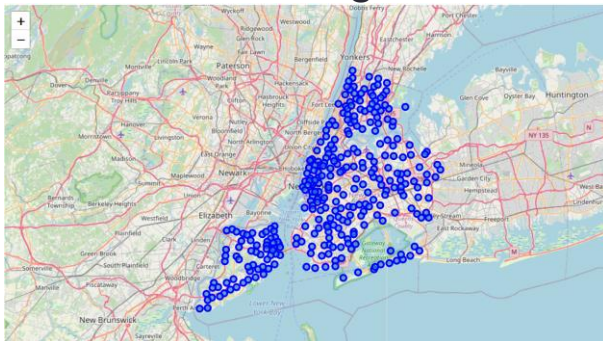
# Data used

- A json file listing neighborhoods and their geographic coordinates
  - [https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/newyork\\_data.json](https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/newyork_data.json)
- Data on the population from <https://geodacenter.github.io/data-and-lab/NYC-Nhood-ACS-2008-12/>
  - [https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/NYC\\_Nhood%20ACS2008\\_12.dbf](https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/NYC_Nhood%20ACS2008_12.dbf)
- NTA code equivalence from [https://www1.nyc.gov/assets/planning/download/office/data-maps/nyc-population/census2010/nyc2010census\\_tabulation\\_equiv.xlsx](https://www1.nyc.gov/assets/planning/download/office/data-maps/nyc-population/census2010/nyc2010census_tabulation_equiv.xlsx)
  - [https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/nyc2010census\\_tabulation\\_equiv.xlsx](https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/nyc2010census_tabulation_equiv.xlsx)
- The study is gathered in this jupyter notebook : <https://github.com/YannPerrin/Applied-Data-Science-Capstone-YP/blob/main/The%20Battle%20of%20Neighborhoods%20YP.ipynb>

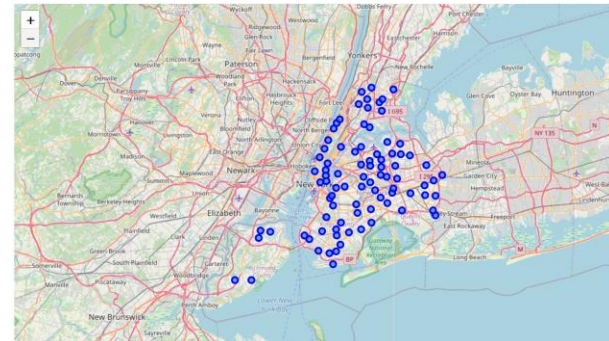
# Data cleaning

- Data cleaning was about merging geographic coordinates data with population data
- In the process, we note a loss of neighborhoods :
  - 306 neighborhood initially
  - 88 neighborhood with comprehensive data
- That means a loss of information, and at the end possibly missing some interesting propositions

All 306 NYC neighborhoods

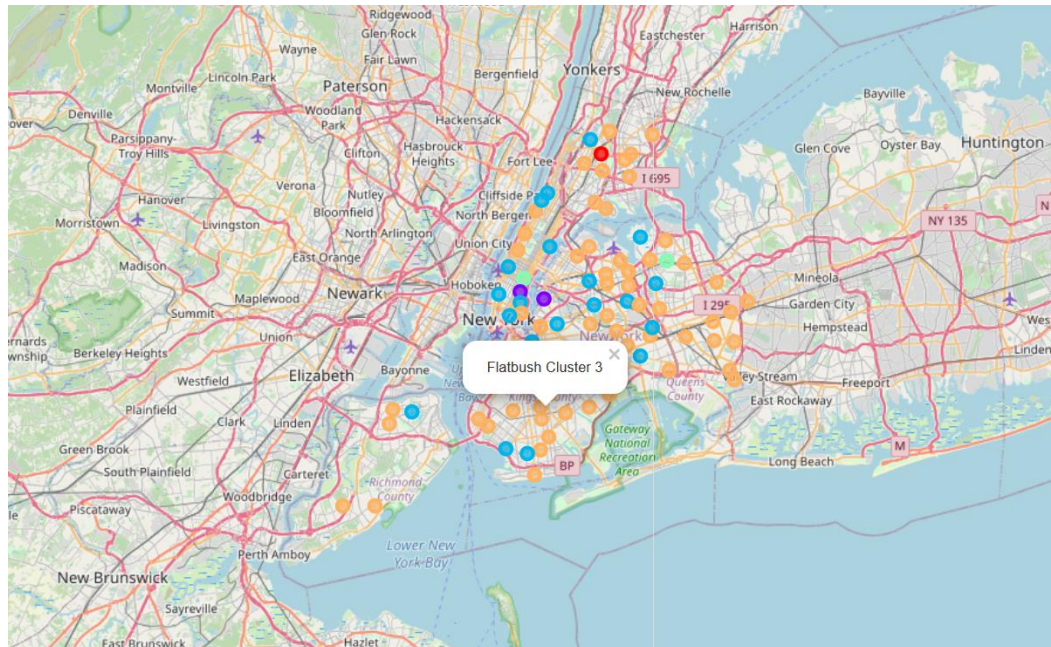


Neighborhoods with comprehensive data



# Clustering similar neighborhoods

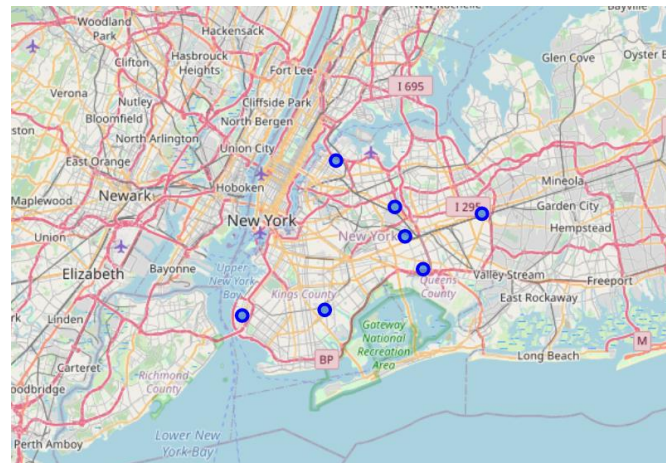
- We consider the venues in a 1000 m radius around the neighborhood location
- Through Kmeans clustering method, we can cluster the neighborhoods in 5 groups based on the number of venues per million inhabitants for each neighborhood
- That shows in orange the neighborhoods similar to Flatbush



# Additional criterias

- From the list of similar neighborhoods, we remove the ones that do not meet the following criterias
  - Have a higher median income than Flatbush → to represent higher life standard
  - Have fewer coffee shop per million inhabitants than in Flatbush → to represent a deficit of coffee shops
- That gives the following list of neighborhoods

Richmond Hill
Queens Village
South Ozone
Park
Flatlands
Bay Ridge
Forest Hills
Astoria



# Conclusion

- Based on data analysis and clustering method, we can identify a list of 7 neighborhoods where a new coffee shop could be open
- To go further, we should understand why in 6 of those 7 neighborhoods there is no coffee shop
  - Does it signal a new coffee shop would meet great success ?
  - Or has it been coffee shops before and that didn't work ?
- Having no further explanation so far, we can suggest Astoria in the Queens to be a good choice :
  - Similar to Flatbush in terms of venues
  - Meeting the two other criterias
  - Already has coffee shops, but with smaller density than in Flatbush
  - Is in another borough