The Battle of Neighborhoods – Week 1

By Hupfstrudel

Assignment:

Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve.

Introduction / Problem:

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

With an estimated population of 8.39 million in an area of 302 square miles, New York City is the most densely populated city in the United States. The City is split into five boroughs (Brooklyn, Queens, Manhattan, Bronx) with a total of 306 neighborhoods. Furthermore, with an overall GMP of 2 trillion US dollars provides endless avenues for potential businesses. However, most of the markets are fairly saturated and as such “bloody”.

Let’s assume a large corporation wants to open a new restaurant in New York City. What types of restaurants are not already oversaturated (e.g. Chinese vs. Indian vs. Italian)? What would be the optimal location for a specific type of restaurant given the demographic of each borough?

Data:

Describe the data that you will be using to solve the problem or execute your idea. Remember that you will need to use the Foursquare location data to solve the problem or execute your idea. You can absolutely use other datasets in combination with the Foursquare location data. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using, even if it is only Foursquare location data.

1. **Datasource:** https://opendata.cityofnewyork.us/
   From „NYC OpenData“ we are able to obtain various data regarding, population, crime rates, health inspections and zip code boundaries

2. **Datasource:** https://geo.nyu.edu/catalog/nyu_2451_34572
   Locational data about the individual boroughs and neighborhoods in New York City.

3. **Datasource:** Geopy
   Just as in the previous assignments we can utilize the geopy package to obtain the latitude and longitude values of New York City and the 5 boroughs.

4. **Datasource:** Foursquare API
   The Foursquare API is utilized to obtain the locations and types of all restaurants in New York City

With these 4 datasources at hand we can first put together a descriptive analysis of the individual boroughs and the restaurant market in general. In the following step we will identify our best option of type of restaurant and corresponding borough. Finally, we can utilize a cluster analysis to group the neighborhoods in the chosen borough to identify the best possible location for the new restaurant.

Needed in Week 2:

Methodology:

Results:

Discussion:

Conclusion: