## Subways

The data for the subway was imported as a shapefile[[1]](#footnote-1) which includes the name of a certain subway station, it’s geometry as a point and other meta data. Then a new table was created containing the distance to the next subway entrance for each NTA zone. This was done by using the postgis ST\_DISTANCE method we have learned during our lectures. At last there was added a rating to the table which represents the normalised distance for each neighbourhood, with ranges from one if there is a subway entrance in the zone to 0 for the zone furthest away from any subway connection[[2]](#footnote-2). Resulting there was a table containing only the id of the NTA and the rating[[3]](#footnote-3) for further processing.

## Soccer fields

The Soccer field data set was imported as a shapefile[[4]](#footnote-4) into the postgis database program. The data relation consists of the geometry of the field as well as of its id and other meta data. Similar to the previously described data processing of the subway, for each NTA area was calculated the distance to the closest football field[[5]](#footnote-5), as there are not that many field in New York. So in the end there was a relation with the id of the NTA and its normalised distance to the next soccer field[[6]](#footnote-6).

## Play areas

The data for the play areas were also imported as a shapefile from the official geo data website of New York City[[7]](#footnote-7). It included the geometry of the park and other meta data which we were of no use for us. First the number of playgrounds in a certain NTA region was calculated by joining the NTA with the playareas table on whether a certain play ground is in that NTA zone using ST\_INTERSECTS. As a further step the same procedure was repeated to find play areas near the zone using ST\_DISTANCE with a distance of 500m, as it seemed reasonable to also considerer those parks. Afterwards those two parts were put together on rating the parks in the NTA area thrice the weight of the ones in the adjacent region[[8]](#footnote-8). Finally, the rating was calculated by considering the total amount of playgrounds in and nearby the zone as well as the zone’s area[[9]](#footnote-9). Those steps were all performed successively as of a lack of performance of the database.

## Parks

The data for the parks and recreational areas[[10]](#footnote-10) of New York City was also downloaded as a shapefile and imported into the postgis database program. The table consists of several columns including the geometry, location, name and other attributes. Firstly, the total area of all parks in a particular NTA zone was calculated by joining the previously imported park table with the NTA one. Therefor the postgis functions ST\_AREA, ST\_INTERSECTS and ST\_INTERSCETION which were presented during our lectures were used. As in this case not only the parks in a certain NTA region, but also the ones close by seem to be relevant, a buffer of 500 m was added around the zone to also determine the possible adjacent recreational areas. Afterwards the total area of all parks counting for a certain NTA region were summed up and divided through the area of the NTA zone to also take that under consideration. For this calculations the database functions ST\_BUFFER, ST\_SETSRID and ST\_DIFFERENCE were used[[11]](#footnote-11). In the end this number was normalised as the rating with a range from zero to one regarding how much park area exists in this NTA neighbourhood[[12]](#footnote-12). These procedures were all executed in small steps as of the performance of the database program.

## Restaurants

To get the data for the restaurants a table[[13]](#footnote-13) was downloaded as a CSV file from the official governmental geo data website of New York City and imported as a text file into the database program. That table included an address with a street name and zip code of a restaurant and several other meta data. As for further processing a geometry data entry was needed to determine the exact location of the restaurant a second data set[[14]](#footnote-14) which matched addresses with a geometry was imported. So, for further processing of that data, irrelevant information like building numbers and other meta data not needed were removed that it contains only valid data with a geometry object and every data only once. After eliminating multiple entries to one restaurant in the restaurant table the two relations were join together using the address with street and zip code. However, this only worked in about half of the cases and the rest could not be matched as of differences in the street name as sometimes there were additionally building numbers in the name or abbreviations were used. As there did not seem to be a better solution the Google Geocoding API was used. For that the not matched restaurants were extracted into another table and exported as text files.

A Java program[[15]](#footnote-15) was developed to read the text files, send for each entry a request to the API which returns the latitude and longitude of the address and save the results in another file. An issue of the API was that the necessity of using a key for the request and the validation of the key which was at about 2500 requests. To avoid this more keys could be requested from the Google API site for other projects. With that solution only about three percentage of the restaurants could not be matched to a location and have to be omitted for further processing.

Afterwards the file with the geo coded data was imported and added to the table with the other restaurants. Finally, the NTA zone’s table was joined with the restaurant’s using the ST\_INTERSECTS function to get the total number of restaurants for each NTA neighbourhood[[16]](#footnote-16). Resulting, the id of the NTA zones and the rating of the total number of restaurants in the particular area were used for further processing[[17]](#footnote-17).

1. https://data.cityofnewyork.us/Transportation/Subway-Stations/arq3-7z49 [↑](#footnote-ref-1)
2. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/subway/subway.sql [↑](#footnote-ref-2)
3. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/subway/subwaydistance\_table.txt [↑](#footnote-ref-3)
4. https://data.cityofnewyork.us/Recreation/Map-of-Soccer-and-Football-Fields/qqsi-vm9f [↑](#footnote-ref-4)
5. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/soccerfields/soccerfields.sql [↑](#footnote-ref-5)
6. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/soccerfields/soccerfield\_data.txt [↑](#footnote-ref-6)
7. https://data.cityofnewyork.us/City-Government/Play-Areas/8fhn-c4v3 [↑](#footnote-ref-7)
8. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/playgrounds/playgrounds.sql [↑](#footnote-ref-8)
9. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/playgrounds/neighbourhood\_playground\_table2.txt [↑](#footnote-ref-9)
10. https://data.cityofnewyork.us/City-Government/Parks-Properties/rjaj-zgq7 [↑](#footnote-ref-10)
11. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/parks/park2.sql [↑](#footnote-ref-11)
12. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/parks/neighbourhood\_parks\_table2.txt [↑](#footnote-ref-12)
13. https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/xx67-kt59 [↑](#footnote-ref-13)
14. https://data.cityofnewyork.us/City-Government/NYC-Address-Points/g6pj-hd8k [↑](#footnote-ref-14)
15. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/restaurants/geoCoding.java [↑](#footnote-ref-15)
16. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/restaurants/restaurants2.sql [↑](#footnote-ref-16)
17. https://github.com/FlorianFusseder/SpatialDatabases/blob/master/data/restaurants/restaurants\_table2.txt [↑](#footnote-ref-17)