

Data

The data for this project will be acquired from two different sources:

1. **Foursquare** will be used to retrieve the list of schools, their geographical locations, and also popular venues near each one.
2. **City-Data.com** will be used to retrieve the list of schools and their rating which is available here: <http://www.city-data.com/schools-dirs/schools-NC.html>

First, the list of schools in Durham will be acquired from Foursquare. Durham is about 110 square miles so I will use a radius of 20 km to be confident that all schools in Durham area will be acquired. Then I will remove schools of other neighboring cities in case there are retrieved by Foursquare and then filter the data for school names and address-related items such as their coordinates in a data frame.

| | name | categories | address | cc | city | country | crossStreet | distance | formattedAddress | labeledLatLngs | lat | |
|---|--|------------------------------|----------------|----|--------|---------------|------------------|----------|---|--|-----------|-------|
| 0 | Merrick-Moore Elementary School | Elementary School | 2319 Cheek Rd | US | Durham | United States | NaN | 2498 | [2319 Cheek Rd, Durham, NC 27704, United States] | [[{"lng": -78.85250003657937, "lat": 36.0051831...}]] | 36.005183 | -78.8 |
| 1 | Club Boulevard Elementary School | Elementary School | W Club Blvd | US | Durham | United States | NaN | 1915 | [W Club Blvd, Durham, NC 27704, United States] | [[{"lng": -78.896396706746, "lat": 36.0190632942...}]] | 36.019063 | -78.8 |
| 2 | North Carolina School of Science and Mathematics | High School | 1219 Broad St | US | Durham | United States | at W Club Blvd | 4153 | [1219 Broad St (at W Club Blvd), Durham, NC 27... | [[{"lng": -78.92128537103233, "lat": 36.0184102...}]] | 36.018410 | -78.9 |
| 3 | Duke Fuqua School Of Business | General College & University | 100 Fuqua Dr | US | Durham | United States | at Science Dr | 6849 | [100 Fuqua Dr (at Science Dr), Durham, NC 2770... | [[{"lng": -78.94743747416524, "lat": 35.9989681...}]] | 35.998968 | -78.9 |
| 4 | Duke University School of Law | Law School | 210 Science Dr | US | Durham | United States | at Towerview Rd. | 6665 | [210 Science Dr (at Towerview Rd.), Durham, NC... | [[{"lng": -78.94568586193236, "lat": 35.9999544...}]] | 35.999954 | -78.9 |

Figure 1. Schools data from Foursquare

Second, city-data.com schools web pages will be parsed for the schools rating data in Durham. The schools of other cities and other irrelevant data will be removed and data will be saved in a data frame.

| | Name | Rating | Grade span |
|---|-------------------------------|--------|------------|
| 0 | Bethesda Elementary | 18.0 | KG-5 |
| 1 | Brogden Middle | 37.0 | 6-8 |
| 2 | Burton Elementary | 21.0 | KG-5 |
| 3 | Carter Community Charter | 22.0 | KG-8 |
| 4 | Central Park School For Child | 74.0 | KG-5 |

Figure 2. Schools rating from city-data.com

The two data frames showing schools names, coordinates, and ranking will be merged and shown on the map using folium. School ranking will be color coded and shown on the map. Then k-means will be used to cluster schools based on their ranking and the clusters will be plotted on the map.

Later, Foursquare will be called to explore the venues in the vicinity of each school and list the top ten popular ones for each school. Finally, a report including the list of public schools in Durham, their address and rating, map of schools and clusters, and top ten venues nearby each school will be prepared and sent to the customer.