**Citibike project**

The goal is to record the number of bikes available every minute for an hour across all of New York City in order to see which station or set of stations is the most active in New York City for that hour. Activity is defined as the total number of bicycles taken out or returned in an hour. So if 2 bikes are brought in and 4 bikes are taken out, that station has an activity level of 6.

**Getting the Data**

Data is located at:

r = requests.get('http://www.citibikenyc.com/stations/json')

root

|\_\_executionTime

|\_\_stationBeanList

|\_\_availableDocks

|\_\_totalDocks

|\_\_city

|\_\_altitude

|\_\_stAddress2

|\_\_longitude

|\_\_lastCommunicationTime

|\_\_postalCode

|\_\_statusValue

|\_\_testStation

|\_\_stAddress1

|\_\_stationName

|\_\_landMark

|\_\_latitude

|\_\_statusKey

|\_\_availableBikes

|\_\_id

|\_\_location

### Challenge

1. Explore the other data variables. Are there any test stations? How many stations are "In Service"? How many are "Not In Service"? Any other interesting variables values that need to be accounted for?
2. What is the mean number of bikes in a station? What is the median? How does this change if we remove the stations that aren't in service?

### Storing Data in database

While often code can be developed iteratively (meaning you create code by making small changes that build on each other), designing a database requires more thoughtful planning, even one as simple as a SQLite database. We know now the structure of the data, which is important for creating the schema we'll be using in database.

Most of these values don't change over time. The only fields likely to change are the availableBikes, availableDocks, statusValue, and the statusKey. In this case, since our goal is to record the number of bikes available every minute for an hour across all of New York City, we're primarily interested in the number of available bikes, but we want to keep all this reference information as well.

Since the number of available docks and available bikes should equal the total docks, we only need to store one of these numbers, in this case, the available bikes. The available docks should be the total docks - the available bikes, but this is a case where we need to be clear about our assumptions and test them if necessary. In the available\_bikes table, we'll be storing the availableBikes value along with the executionTime.

**Challenge**

1. Now that the tables have been created, create a script based on the code you've written so far that downloads the data, parses the result, and then uploads the data to the database.
2. The code then needs to sleep for a minute and then perform the same task. Find a way to make this happen in your code.
3. The code only needs to run for an hour. If it's sleeping every minute, the code only needs to loop 60 times. Find a way of doing this
4. Construct an algorithm (writing code) to see which station is the most active in New York City in that hour.