We took the GeoText2010 twitter database which contains tweets by the users in USA.

The columns present in the dataset are – twitter handle of the user, the latitude and longitude of the place where user created the account, the tweet text of all the tweets by the users. The total number of users were around 9500 and total number of tweets were approximately 850,000.

The aim was to take the latitudes and longitudes of all the users and map it to a particular state and region of USA. For this purpose, we used the reverse geocoding technique. We used a free Mapquest API which gives all the location address attributes of a lat-long combination in a JSON format.

The tasks were as follows -

1. Convert the given TSV (Tab Separated Values) file to CSV. We wrote a python script for this purpose.
2. Remove the unwanted columns from CSV. We wrote a python script for this purpose.
3. Read the above CSV, call the API for every lat-long value, parse the results and store it in a CSV file. We wrote a JavaScript code for this purpose which runs on Nodejs server.
4. Combine the results with users and tweets text. We wrote a python script for this purpose.