* Method1. Download the data from NYC open data and put it on HDFS

<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>

\*\* we can get the data from 2010 -2019

* Method2. Using Soda API ( so we can update the data weekly or monthly)

python get311dataclean.py 1000000 2019-03-01 2019-04-01 Mar2019.csv Mar2019\_clean.csv