# Boston 311 Project README

Jane Huang, Isadora Nun, Weiwei Pan, Francisco Rivera

## 1 Project Files

1. `Code` Directory
   (a) File `Reading in Data.ipynb`: reads the 311 data into a `pandas` data frame. Supports basic exploration of the entire set of 311 data.
   (b) File `Basic Analysis of Data.ipynb`: provides rudimentary visualization and analysis of closed requests generated in the year of 2015. Contrast and comparison between requests originating from call data and Citizens Connect App data is emphasized throughout.
   (c) File `Simulated Annealing.ipynb`: provides maximum likelihood estimations of the mixture model parameters through simulated annealing. Includes convergence analysis and visualizations of hard clusterings of the data based on mixture parameter estimates. Contrast and comparison between requests originating from call data and Citizens Connect App data is emphasized throughout.
   (d) File `EM for MLE and MAP.ipynb`: provides maximum likelihood and maximum a posteriori estimates of the mixture model parameters through expectation maximization. Includes:
   - performance testing on synthetic data
   - model selection for the number of mixture components using Bayesian information criterion
   - convergence analysis on real data
   - hard clustering of the data based on MAP mixture parameter estimates
   - cluster profile analysis

   Contrast and comparison between requests originating from call data and Citizens Connect App data is emphasized throughout.
   (e) File `Gibbs Sampler for GMM.ipynb`: provides Gibbs sampling from the posterior distribution of the mixture model. Include:
   - convergence analysis
   - hard clustering of the data based on posterior mean estimates of the mixture parameters
   - visualization of the posterior predictive
   - comparison of performance against basic MH sampling implemented in `PyMC`
   - alternative 1-D model for response time as a mixture of exponentials

   Contrast and comparison between requests originating from call data and Citizens Connect App data is emphasized throughout.
2. `Paper` Directory
   (a) File `Boston311_poster.pdf`: poster with overview of the project including select results
   (b) File `Boston311_paper.pdf`: summary, with details, of the methods and results as well as analysis.
3. File `Boston311_screen_cast.mp4`: 2 minute screen-cast with brief overview of the project including select results.

## 2 Downloading the Data

The Boston 311 dataset (`.csv` formatting required) can be downloaded at

https://data.cityofboston.gov/City-Services/311-Service-Requests/awu8-dc52

The data is `360MB` in size.

# 3 Running the `ipython` Notebooks

Each `ipython` notebook can be run independently and each requires initially reading and processing of the 311 dataset (a process requiring approximately `300` seconds). The notebooks include section headers, expository narrative and comments throughout to facilitate exploration.