Data Mechanics
Assignment 1

We proceeded with 3 transformations combining data from 3 different portals as suggested by the description of the assignment.

In conjunction with our Spark projects, we used ballot questions, primary and general elections data as well as demographics data concerning towns,cities, districts etc.

Transformation 1:

Gathering ballot questions answers and calculating averages after aggregating by town. Data included results for yes, no, blank and total votes for each district, ward, town of Massachuseetes.

Data is loaded and stored in mongo at questions.py. Aggregations and the store of the final result happens at questionsAggr.py. The outcome of this transformation is an average number of people who answered the questions for each column. This is an example of how we will use the ballot questions further as part of our Spark project.

Transformation 2:

Collect data of democratic primary election, republican primary election and green-rainbow primary election and compute the total number of voters for each stage respect to counties in Massachusetts. The output includes three collections and each of which has county as key and total number of voters as value for each record.

This transformation is implemented by two scripts where fetch.py fetches all concerned datasets online and saves them into mongo and transformTotalVoter.py does several built-in functions then. The outcome allows us to invest the inclination of folks to vote according to political preferences and geographical difference.

Transformation 3:

This transformation gathers voting statistics on Massachusetts counties (including blank ballots) from the general election and aggregates them despite ward and individual areas within each county. The data is loaded and stored in MongoDB from generalElection.py. The result of the transformation will allow us to more easily map out voter statistics by location, which is why we maintained the keys (counties) from the original data.