My cleaning file:

http://pastebin.com/0qN3DZ2d

My Mongodb queries

http://pastebin.com/uWqgqFQz