Comments on dataset and processing

Dataset is the file *requetes311.csv* (text contains weird symbols due to French language)

Dataset has three data fields:

ACTION\_NOM – type of request/information

DDS\_DATE\_CREATION – date of request/information

UNITE\_RESP\_PARENT – district of Montreal from where information/request was done

Processing of dataset using different MapReduce algorithms can provide information on:

-the most frequent request

-total number of requests per each district of city

-total number of requests per day

-the most frequent request per each district

-etc.

I chose an option “the most frequent request per district” (see file MapReduce1\_Korchagina.py). and “total number of requests per each district of city” (see file MapReduce2\_Korchagina.py).