REGISTER /usr/lib/pig/piggybank.jar;

extract\_details = LOAD '/user/cloudera/data/AD1.csv' using org.apache.pig.piggybank.storage.CSVLoader() AS (id:long,text:chararray);

tokens = foreach extract\_details generate id,text, FLATTEN(TOKENIZE(text)) As word;

dictionary = load '/user/cloudera/data/AFINN.txt' using PigStorage('\t') AS(word:chararray,rating:int);

word\_rating = join tokens by word left outer, dictionary by word using 'replicated';

describe word\_rating;

rating = foreach word\_rating generate tokens::id as id,tokens::text as text, dictionary::rating as rate;

word\_group = group rating by (id,text);

avg\_rate = foreach word\_group generate group, AVG(rating.rate) as tweet\_rating;

STORE avg\_rate INTO 'SentimentAnalysis' using PigStorage(',');