business\_data =

LOAD 'yelp\_academic\_data\_business.json'

USING JsonLoader(

'business\_id:chararray,

name:chararray,

neighborhood:chararray,

address:chararray,

city:chararray,

state:chararray,

postal\_code:chararray,

latitude:float,

longitude:float,

stars:float,

review\_count:int,

is\_open:int,

attributes:bag{a:tuple(a:chararray)},

categories:bag{a:tuple(a:chararray)},

hours:bag{a:tuple(a:chararray)},

type:chararray'

);

business = FOREACH business\_data GENERATE city, review\_count, FLATTEN(categories) as category;

groups = GROUP business BY (city, category);

result = FOREACH groups GENERATE FLATTEN(group) AS (city, category), SUM(business.review\_count);

ordered = ORDER result BY city;

STORE ordered INTO 'yelp/ans1' USING PigStorage(',');