

BizStar ReadMe

Introduction:

This part gives average star of businesses in all locations. To give an objective evaluation of prosperity of one location. For every business in a specific location, the business's star is multiplied with review count of the business. And all these multiplication products are added together and then divided by the sum of all review count in that location. The calculation formula is as the following:

$$\text{AvgStar} = \frac{\sum_{\text{businesses } b \text{ within the location}} b.\text{reviewcount} * b.\text{star}}{\sum_{\text{businesses } b \text{ within the location}} \text{reviewcount}}$$

Mapper: In Mapper class, the output key is location number.

The value includes two values separated by a tab that write in one string. Because not only the product of star and review count is needed, the total review count is also needed in Reduce class. So the output value is

`<business.star*business.reviewcount + "\t" + business.reviewcount >`

Output of Mapper class is `<key: location number, value: phrase>`

Reducer: In Reducer class, I add all `business.star*business.reviewcount` from Mapper class together and I also add all `business.reviewcount` together. And the output key of the Reducer class is location number, the value is:

$$\frac{\text{SUM}(\text{business.star} * \text{business.reviewcount})}{\text{SUM}(\text{business.reviewcount})}$$

Input: bizstarinput_clean.tsv (This file is from the line 37-38 in HiveRelatedCode in HiveRelated folder. And this input is actually the same as ngraminput_clean.tsv)

Output: bizstarOutput.tsv (The following is the result screenshot)

```
1 100-2.9831286486991573
2 105-5.0
3 113-3.0
4 114-3.554165871804655
5 115-4.0
6 116-3.192982456140351
7 117-3.157283288650581
8 118-3.9963963033713665
9 119-3.906993006993007
10 120-3.093152477839253
11 13-3.5
12 132-1.2
13 133-4.974763406940063
14 134-3.2099869167030093
15 135-3.801324503311258
16 136-3.898038321167883
17 137-4.083795782463929
18 138-3.087525954320396
19 139-3.7444802388622613
20 140-3.395654752904779
21 146-4.5
22 147-1.5
23 149-3.5
24 15-3.550338091660406
25 150-4.5
26 152-1.0
27 153-1.8762886597938144
28 154-4.002187231078146
29 155-3.999571701458785
30 156-1.937810573068836
31 157-3.555029585798817
32 158-2.9292035398230087
33 159-2.84626890025019
```

