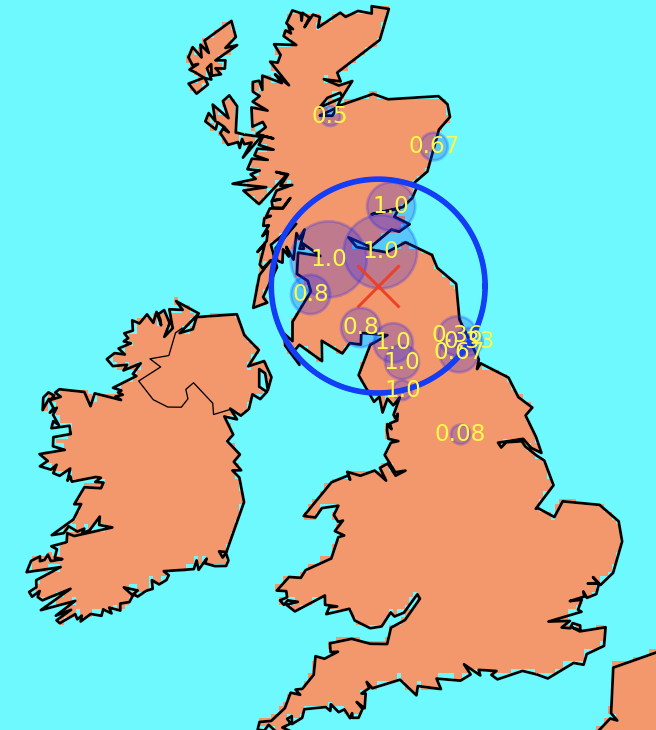
The project finds that it is insufficient to use only the radius of the word range as metrics. For example, the radius of ‘balti’ (Fig. 4) is large, but most of the words are concentrated in a small area. Thus, the project tries to use more features (central coordinate, median of distance from the central coordinate, the number of shops whose distance from the central point less than the median distance, the number of shops whose distance from the central point larger than the median distance, the largest distance of the word from the central point and the average distance of the word from the central point) to better judge the regionality of a word. The project selected 16 words and each of them distributed in more than 4 cities. In these 16 words, 8 of them are localized words and 8 are widely-distributed words. The project wants to find commonalities and differences between regional words and non-regional. Besides, the project selected 8 words which are distributed in 2 or 3 cities to explore and their metrics which are used for making judgement. However, in terms of words which only distributed in one city, the project cannot use above features to classify them and the project wang to initially mark them as unknown.

**Regional words findings (each of them distributed in more than 4 cities):**

Through comparing 8 regional words, the project has found some regionality words’ commonalities:

1. The number of cities is less than the half of total city number: About 15 or less.
2. The number of shops whose distance from the central point less than the median distance/ The total number of shops contain this word > 70% (less/total>70%).
3. Although there are some of words’ radius is large, such as the ‘balti’, the average distance is all lower than 300000.

****

**Figure 1: Haggis distribution range (95% shops)**

lower than half cities:15.

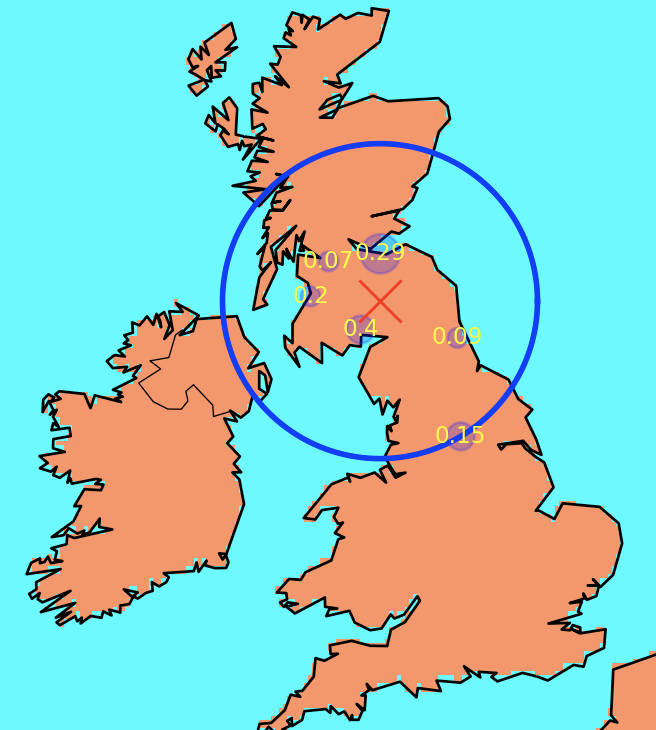
radius: 244021.64.

average distance: 216926.14.

median< average distance and radius.

the largest distance >Radius > average distance.

median (less/total>70%).



**Figure 2: Quattro distribution range (95% shops)**

very few cities: 6.

radius: 359450.13.

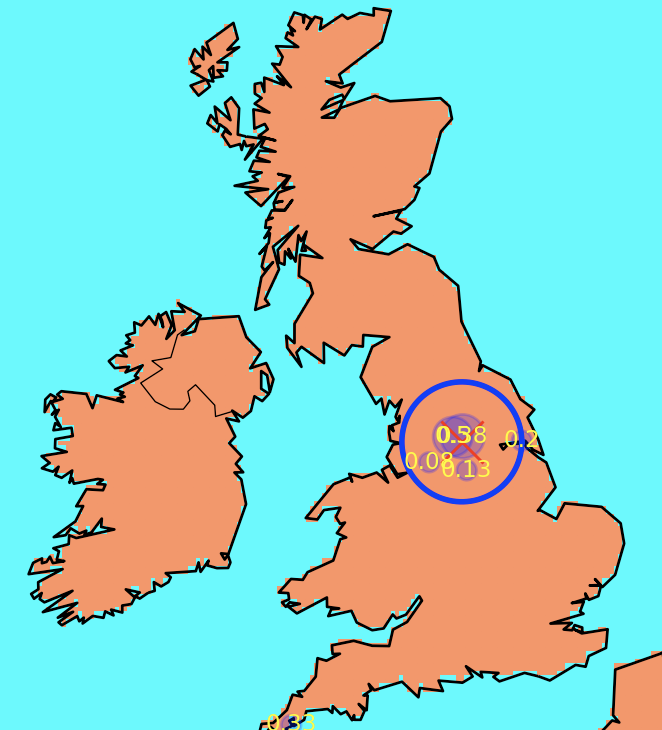
average distance: 174770.13.

median< average distance and radius.

radius not small.

largest distance=Radius > average distance.

median (less/total>70%).

****

**Figure 3: Yorkshire distribution range (95% shops)**

very few cities: 6.

radius: 136824.84.

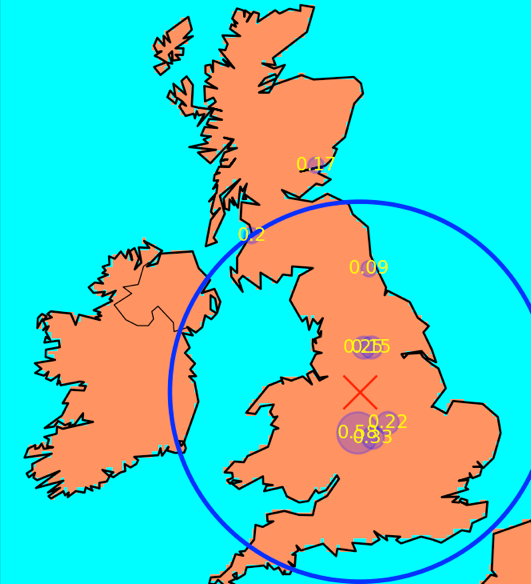
average distance: 180269.94.

median< average distance and radius.

radius is small.

radius < average distance (5000).

median (less/total>70%).



**Figure 4: balti distribution range (95% shops)**

very few cities: 8.

radius: 542117.52.

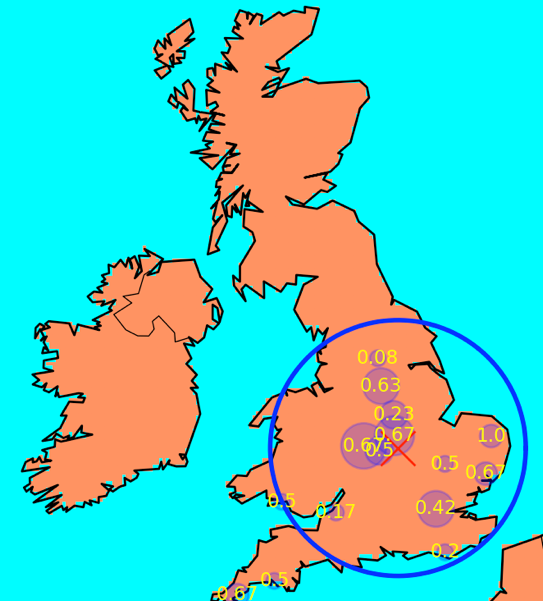
average distance: 272809.68.

median< average distance and radius.

radius not small.

the largest distance >Radius > average distance.

use median (less/total>70%) to judge.



**Figure 5: roe distribution range (95% shops)**

lower than half cities:15.

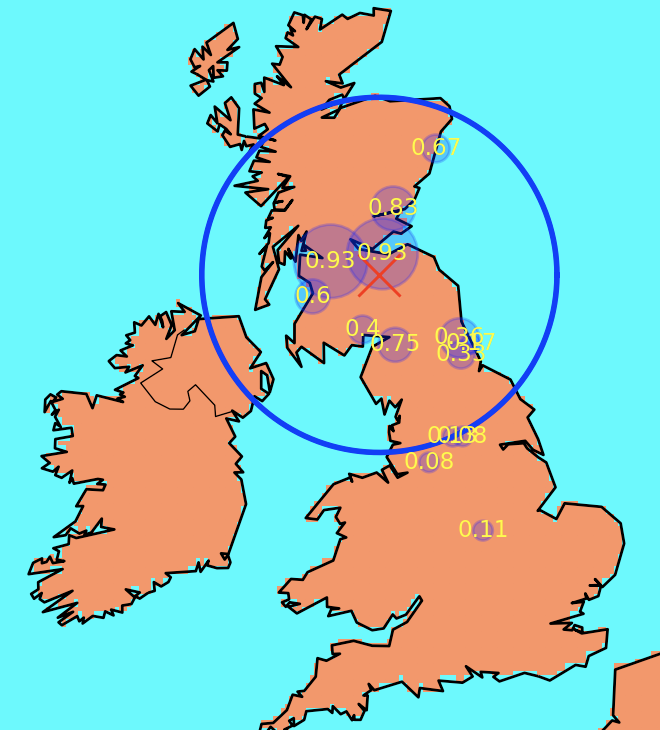
radius: 355658.54.

average distance: 239904.54.

median< average distance and radius.

the largest distance >Radius > average distance.

median (less/total>70%).

****

**Figure 6: irn distribution range (95% shops)**

Irn-Bru Scottish carbonated soft drink.

lower than half cities:14.

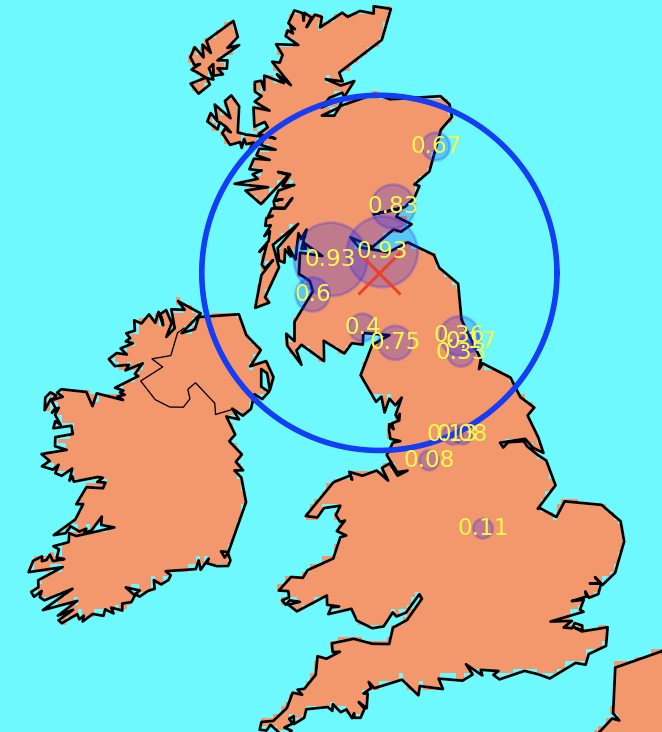
radius: 405285.07.

average distance: 266813.33.

median< average distance and radius.

the largest distance >Radius > average distance.

median (less/total>70%).

****

**Figure 7: bru distribution range (95% shops)**

lower than half cities:14.

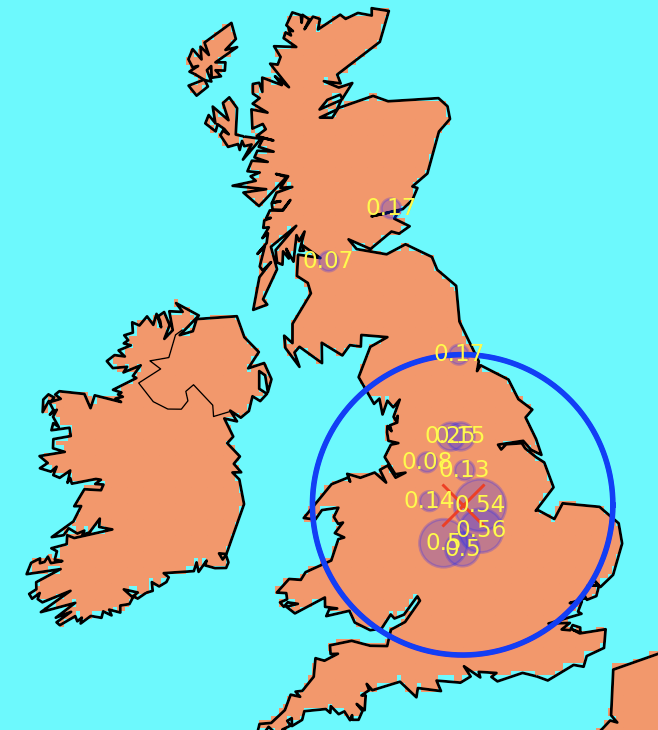
radius: 405285.07.

average distance: 266813.33.

median< average distance and radius.

the largest distance >Radius > average distance.

median (less/total>70%).

****

**Figure 8: naan distribution range (95% shops)**

lower than half cities:12.

radius: 342806.84.

average distance: 214933.34.

median< average distance and radius.

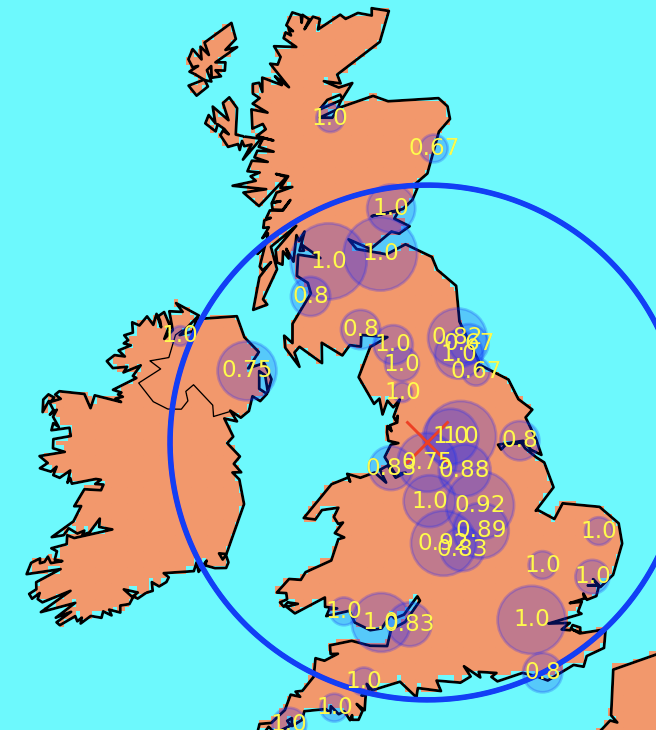
the largest distance >Radius > average distance.

median (less/total>70%).

**Widely distributed words findings (each of them distributed in more than 4 cities):**

Through comparing 8 widely distributed words, the project has found some widely distributed words’ commonalities:

1. The number of cities is more than the regional words have: Among the existing discoveries, words that account for more than half of the total number of cities are widely distributed. However, there are also widely distributed words which have less than half of the total number of cities.
2. The number of shops whose distance from the central point less than the median distance/ The total number of shops contain this word < 70% (less/total<70%).
3. In these 8 widely-distributed words, the radiuses of these words are greater than the radiuses of the above-mentioned regional words.
4. The average distance is greater than 300,000 and greater than the average distance of the above-mentioned regional words.

****

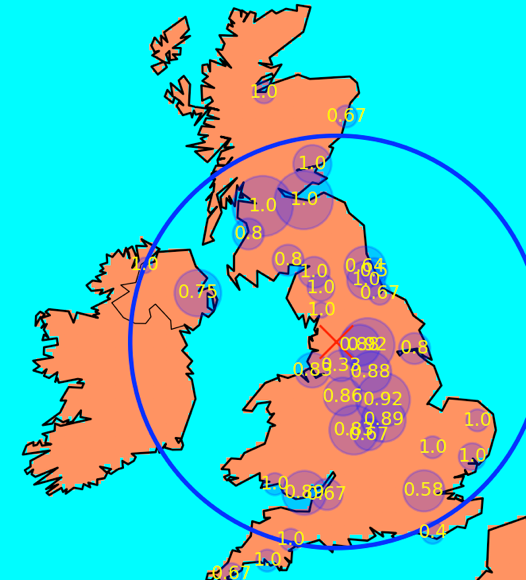
**Figure 9: chips distribution range (95% shops)**

all cities contain this word: 38 cities.

radius= 587407.29.

average distance= 351005.86.

median (less/total=55%<70%).



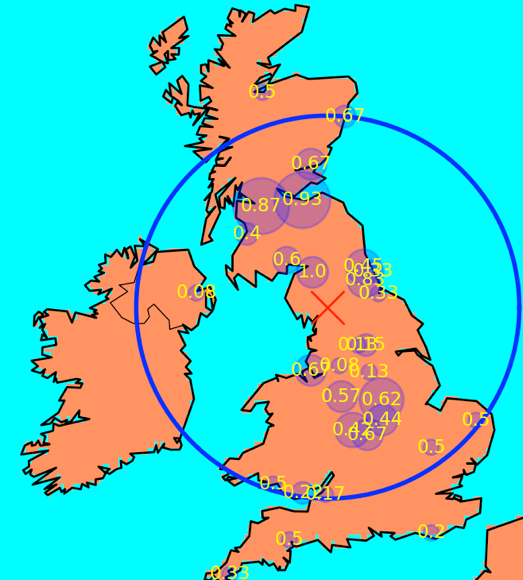
**Figure 10: sausage distribution range (95% shops)**

all cities contain this word: 38 cities.

radius= 592374.47.

average distance= 353703.24.

median (less/total=55%<70%).



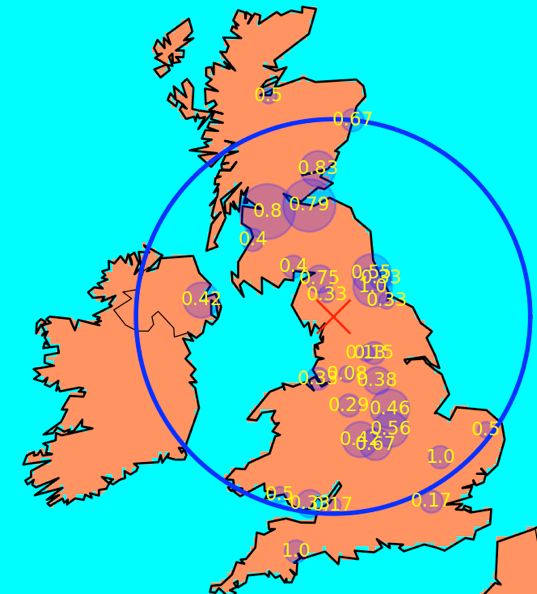
**Figure 11: ham distribution range (95% shops)**

most of cities contain this word: 31 cities.

radius= 550150.25.

average distance= 363073.2.

median (less/total=56%<70%).



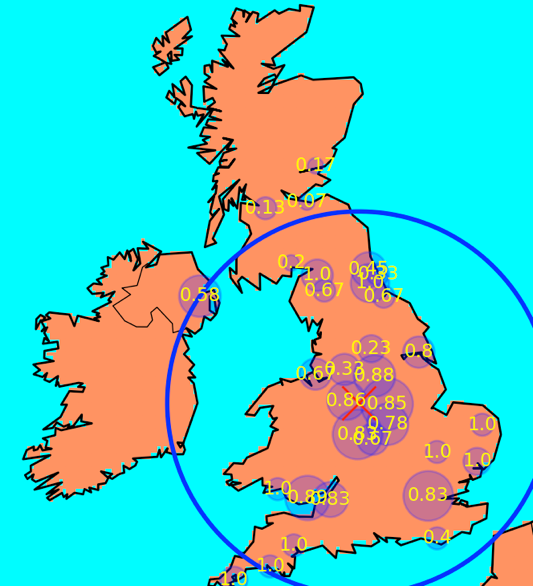
**Figure 12: fanta distribution range (95% shops)**

most of cities contain this word: 31 cities.

radius= 556432.61.

average distance= 334651.6.

median (less/total=39%<70%).



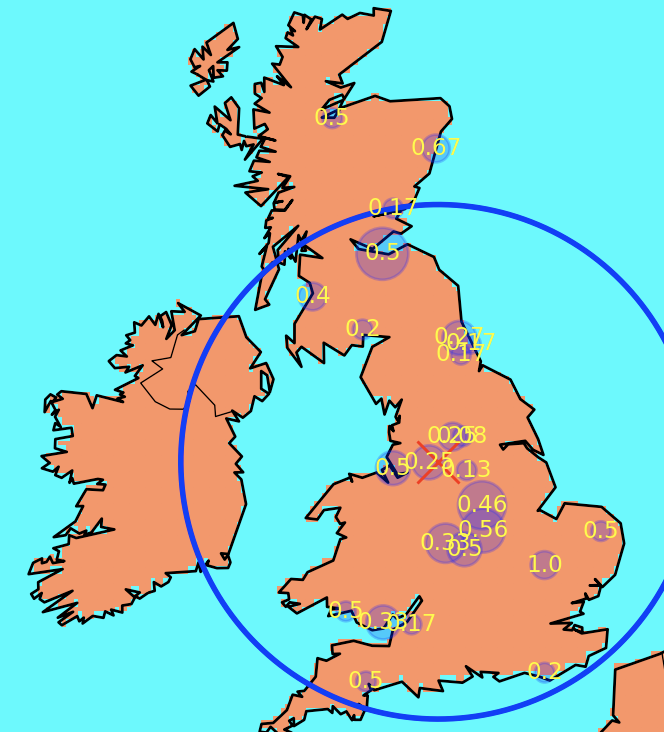
**Figure 13: cod distribution range (95% shops)**

most of cities contain this word: 32 cities.

radius= 545044.82.

average distance= 325843.

median (less/total=66%<70%).



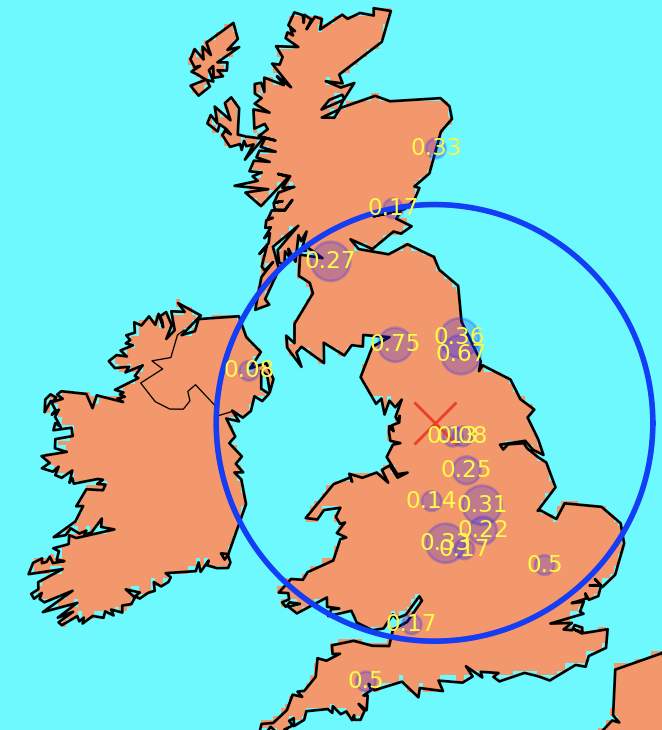
**Figure 14: lamb distribution range (95% shops)**

more than half cities contain this word: 25 cities.

radius= 587105.42.

average distance= 331538.89.

median (less/total=57%<70%).



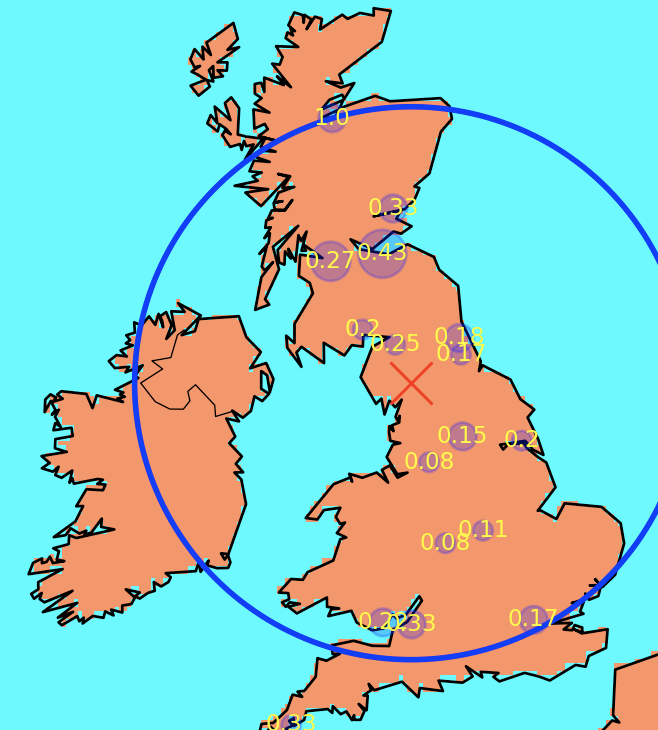
**Figure 15: supreme distribution range (95% shops)**

lower than half 15 cities contain this word: 18 cities.

radius= 498237.39.

average distance= 308121.37.

median (less/total=59%<70%).



**Figure 16: cakes distribution range (95% shops)**

lower than half cities contain this word: 17 cities.

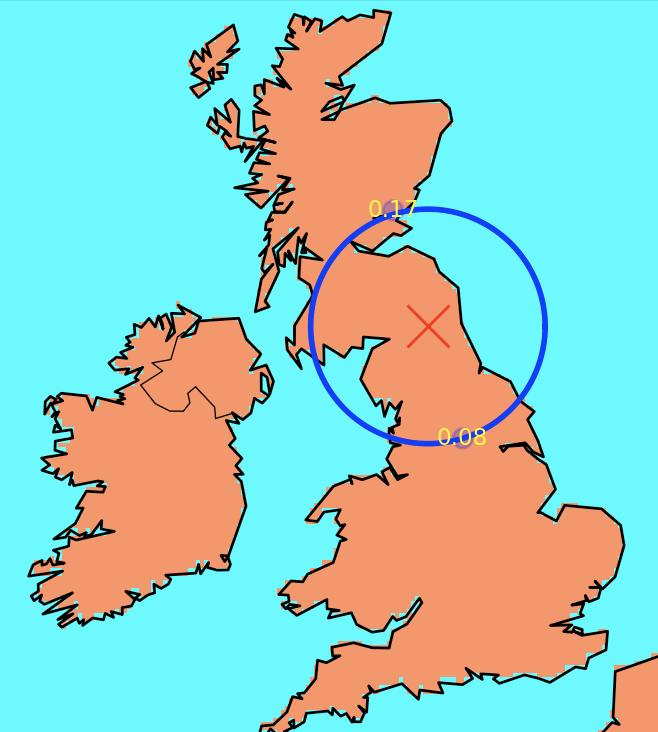
radius= 630944.91.

average distance= 360647.19

median (less/total=46%<70%).

**Words which distributed in 2 or 3 cities findings:**

Through observing the data set, the project selected 8 words that are relatively strange. The project found that the radius of regional words which only in few cities all lower than 120000. Although there are some non-regional words such as ‘book’ and ‘rest’ within this distance, the project initially decided to use radius<120000 to judge regional words which distributed in 2 or 3 cities.

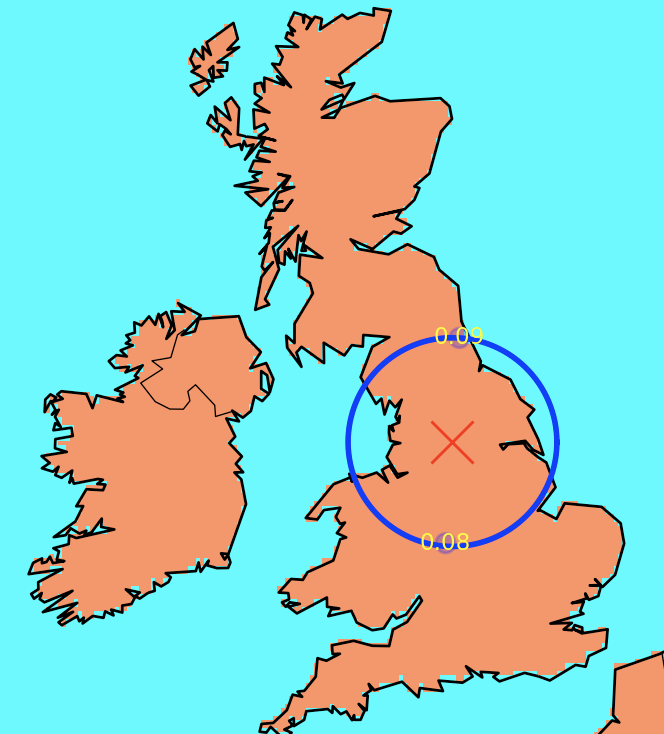


**Figure 17: Dhansak distribution range**

The number of cities: 2

median (less/total=50%).

Radius: 276149.01.

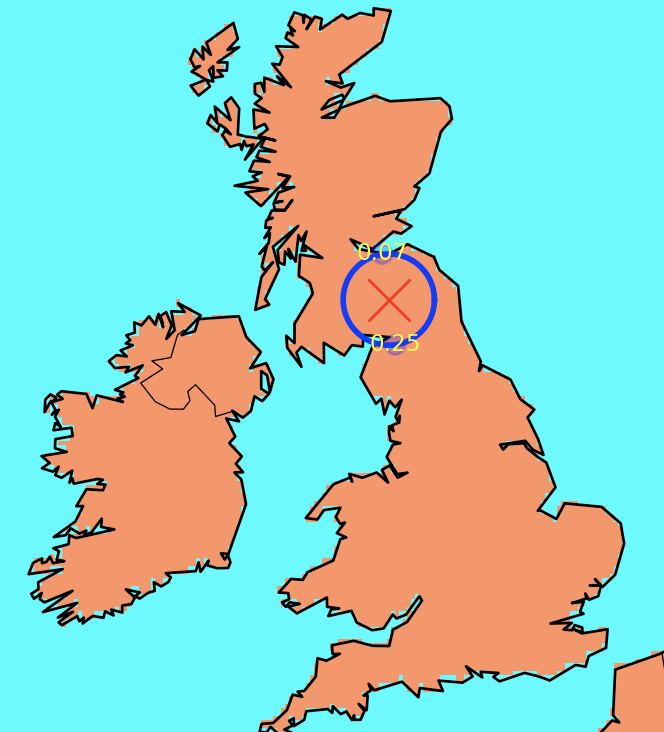


**Figure 18: jalepeno distribution range**

The number of cities: 2

median (less/total=50%).

Radius: 238243.69.

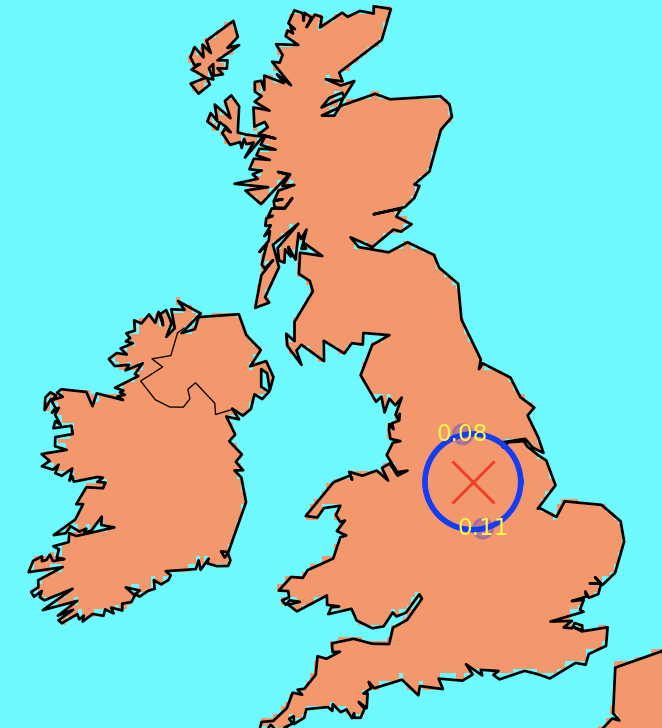


**Figure 19: kopparberg distribution range**

The number of cities: 2

median (less/total=50%).

Radius: 105564.76.

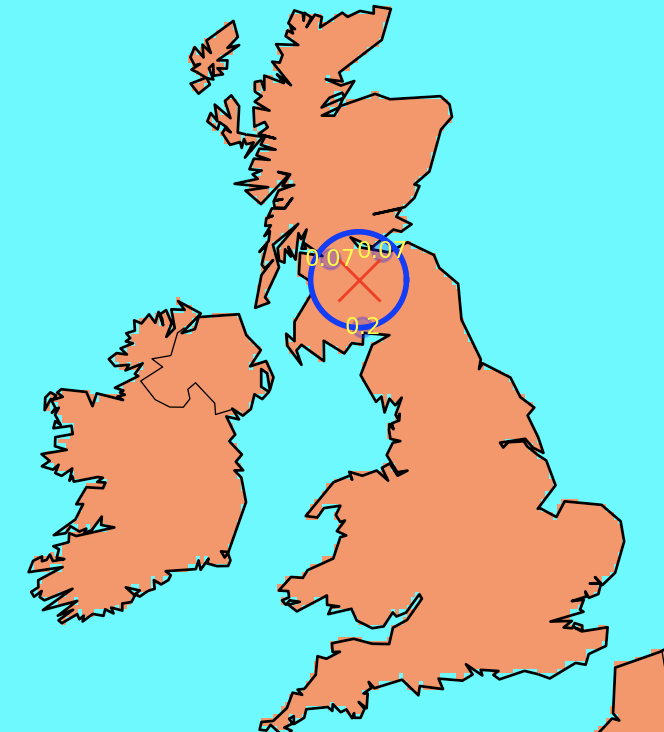


**Figure 20: cabernet distribution range**

The number of cities: 2

median (less/total=50%).

Radius: 110698.34.



**Figure 21: scone distribution range**

The number of cities: 3

median (less/total=33%).

Radius: 109451.98.

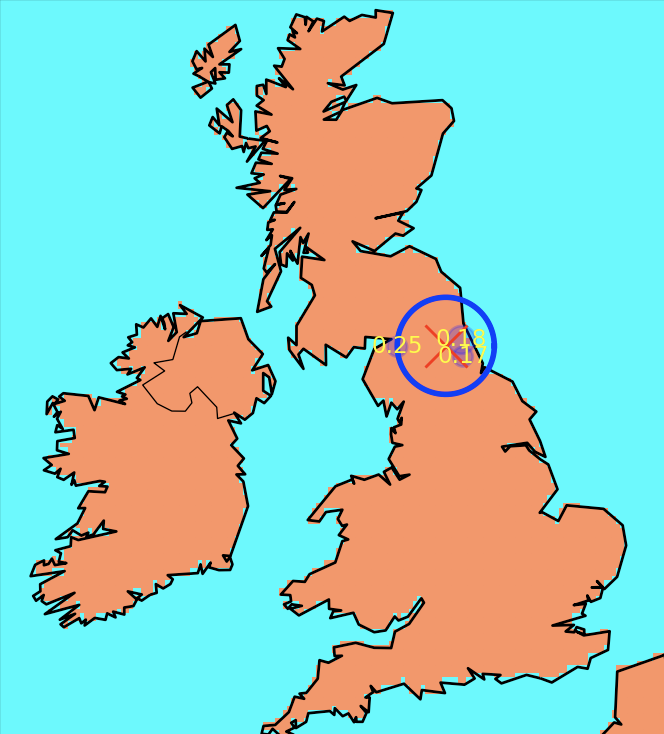


**Figure 22: buns distribution range**

The number of cities: 3

median (less/total=50%).

Radius: 111848.26.

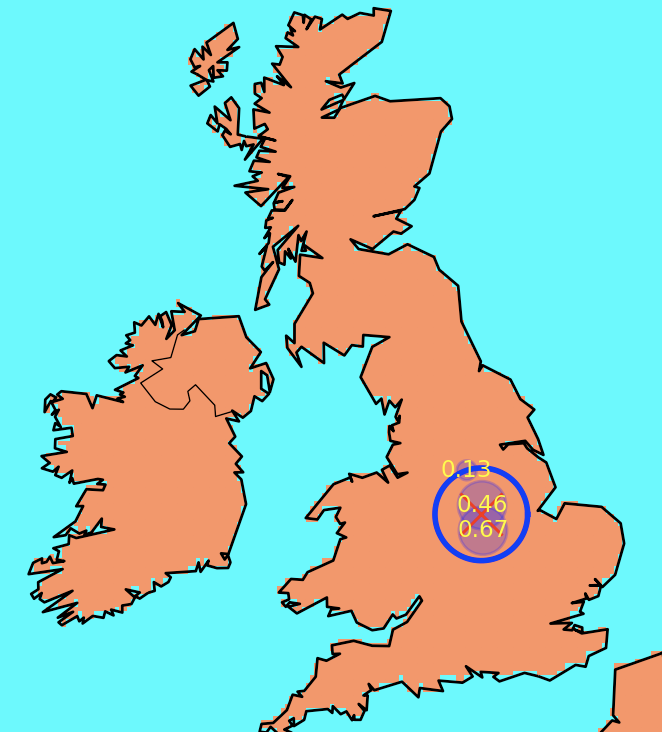


**Figure 23: massala distribution range**

The number of cities: 3

median (less/total=50%).

Radius: 110787.24.



**Figure 23: cob distribution range**

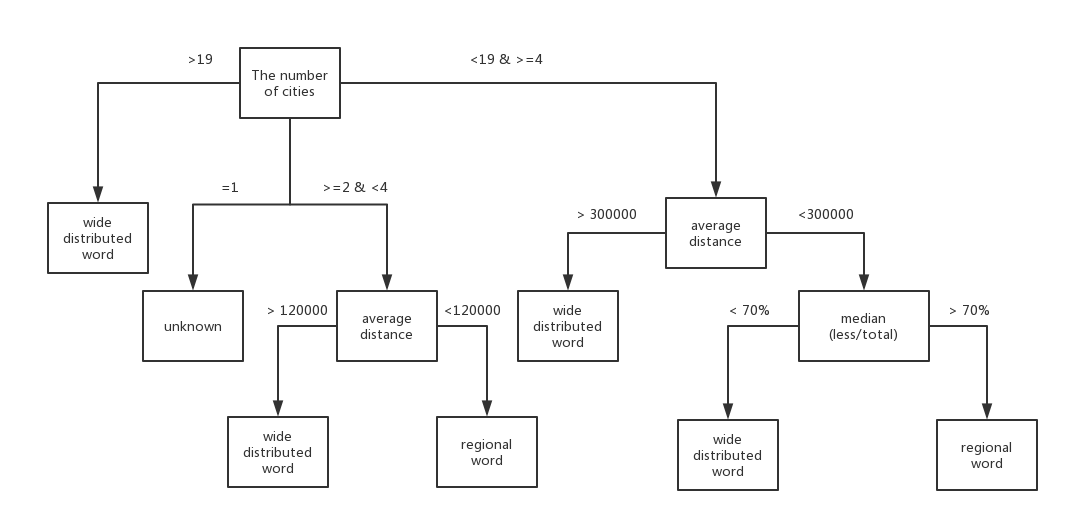
The number of cities: 3

median (less/total=46%).

Radius: 105748.974.

**Conclusion:**

Project plan to create a decision tree (Fig. 24) based on the above judgment method to determine the regional words.



**Figure 24: Judgement decision tree**