**Symbols used in this text:**

**N** number of reviews for each hotel

**TS**  trust score

**W** weight function for trust score

**RTS** real trust score

**C** Constant in the weight function, see detailed definition below

 90% reliable number of review, see detailed definition below

**The Definition of Weight Function**

From the data description, we know that **'Trust Score'** represents the percentage of people who gave positive opinion on the survey. Obviously, with this score alone we cannot determine whether a hotel is good or not, because the numbers of reviews are different for various hotels. The score for a hotel with more reviews should be better to reflect its quality. Thus we need to add a **weight function** based on **the number of reviews** for the calculation of '**Real trust score**'.

There are some requirements for the weight function: **W(N)**

1. **W(N)** should be an **increasing function.** The more reviews there are, the more reliable the trust score is.

2. When the number of reviews **N** goes to infinity, **W(N)** should goes to 1, which means the 'Trust Score' is absolutely reliable.

3. The slope of the weight function should decrease when N increase. The weight increases faster when N is small, slower when N is large. For example, Hotel A scored 90 with 1000 reviews, Hotel B scored 89 with 2000 reviews. If weight is increased linear with N, Hotel B's rank will be significantly higher than Hotel A, which is unreasonable.

A simple function can satisfy the two requirements above:

, C is a positive constant.

when **N**=0, **W(N)**=0; **N** goes to infinity, **W(N)**=1.

The value of **C** should be determined based on the popularity of the location.

To calculate **C**, we can set a 'reliable number ' of reviews. When the number of reviews reach this reliable number, the weight is 90%.



For example, in City San Francisco is a popular choice for tourist. We can set which means the score worth 90% percent of real trust when the hotel get 50 reviews.

In a less popular tourist city, this limit can be lower to a reasonable value. For example, set  for new Berkeley, which means the score worth 90% percent of real trust when the hotel get 20 reviews.

In program, I test to see that, set  to 20 in Berkeley, and 50 in San Francisco will give us a good result.

Then we could compute the average review counts of the top 10% review counts.

Average(top 10%) in Berkeley = 1275

Average(top 10%) in Berkeley = 2796

So, we could have a linear regression on these 2 points(1275, 20), (2796, 50), which gives a function:

**N0.9 = Average(top 10%)/50.7 - 5.15**

In summary, the process I went through is:

1. Compute the average review counts of the top 10% review counts.

2. **N0.9 = Average(top 10%)/50.7 - 5.15**

3. 

4. , N is the review\_counts parameter for a hotel.

5. final\_score = Trust\_Score\*W(N)

In this way, I can only give final score to the hotels that are already been rated by at least one customer. So the hotels with no reviewers will not be ranked.