# Contents



# 1 Each colom represent mean

| Column | Mean | remark |
|---|---|---|
| F (6) | ranking | |
| G (7) | Chinese origin data | |
| H (8) | Google translated data | |
| I (9) | Baidu Chinese sentiment analysis positive probability | base on Chinese origi |
| J (10) | Baidu Chinese sentiment analysis confidence | base on Chinese origi |
| K (11) | Baidu Chinese sentiment analysis Negative probability | base on Chinese origi |
| L (12) | Baidu Chinese sentiment analysis the category | base on Chinese origi |
| M (13) | Baidu English sentiment analysis positive probability | base on Google transl |
| N (14) | Baidu English sentiment analysis confidence | base on Google transl |
| O (15) | Baidu English sentiment analysis Negative probability | base on Google transl |
| P (16) | Baidu English sentiment analysis the category | base on Google transl |
| Q (17) | Google Chinese sentiment analysis score | base on Chinese origi |
| R (18) | Google Chinese sentiment analysis manitude | base on Chinese origi |
| S (19) | Google English sentiment analysis score | base on Google transl |
| T (20) | Google English sentiment analysis manitude | base on Google transl |
| U (21) | Yandex translated data | base on Chinese origi |
| V (22) | Google English sentiment analysis score | base on Yandex trans |
| W (23) | Google English sentiment analysis manitude | base on Yandex trans |
| X (24) | Baidu translated data | |
| Y (25) | Google English sentiment analysis score | base on Baidu transla |
| Z (26) | Google English sentiment analysis manitude | base on Baidu transla |
| AA (27) | Baidu English sentiment analysis postitive probability | base on baidu transla |
| AB (28) | Baidu English sentiment analysis confidence | base on baidu transla |
| AC (29) | Baidu English sentiment analysis Negative probability | base on baidu transla |
| AD (30) | Baidu English sentiment analysis the category | base on baidu transla |
| AE (31) | Baidu postitive probability change to Google score standard | base on origin data B |
| AF (32) | Baidu postitive probability tranform to Google score standard | base on column M |

## 1.1 note

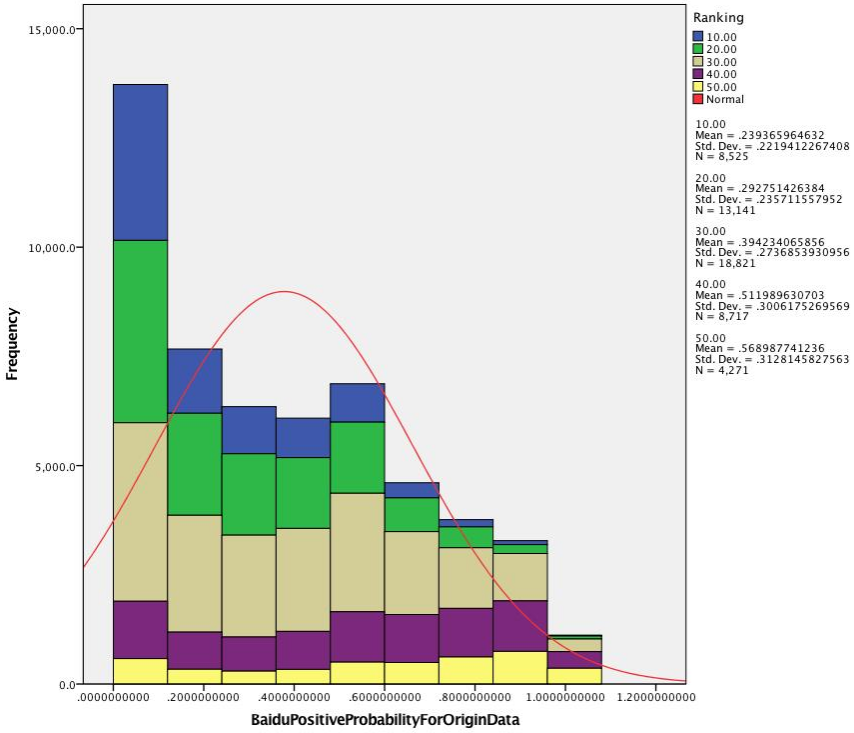
- Baidu sentiment analysis Category Note
  - 2 mean to belong to the positive category, 1 mean to belong to the neutral category, and 0 mean to belong to the Negative category
- Google sentiment analysis Note
  - The score of a document's sentiment indicates the overall emo-

tion of a document. The magnitude of a document's sentiment indicates how much emotional content is present within the document, and this value is often proportional to the length of the document.

- A document with a neutral score (around 0.0) may indicate a low-emotion document, or may indicate mixed emotions, with both high positive and negative values which cancel each out. Generally, you can use magnitude values to disambiguate these cases, as truly neutral documents will have a low magnitude value, while mixed documents will have higher magnitude values.
- Clearly positive and clearly negative sentiment varies for different use cases and customers. You might find differing results for your specific scenario. We recommend that you define a threshold that works for you, and then adjust the threshold after testing and verifying the results. For example, you may define a threshold of any score over 0.25 as clearly positive, and then modify the score threshold to 0.15 after reviewing your data and results and finding that scores from 0.15-0.25 should be considered positive as well.

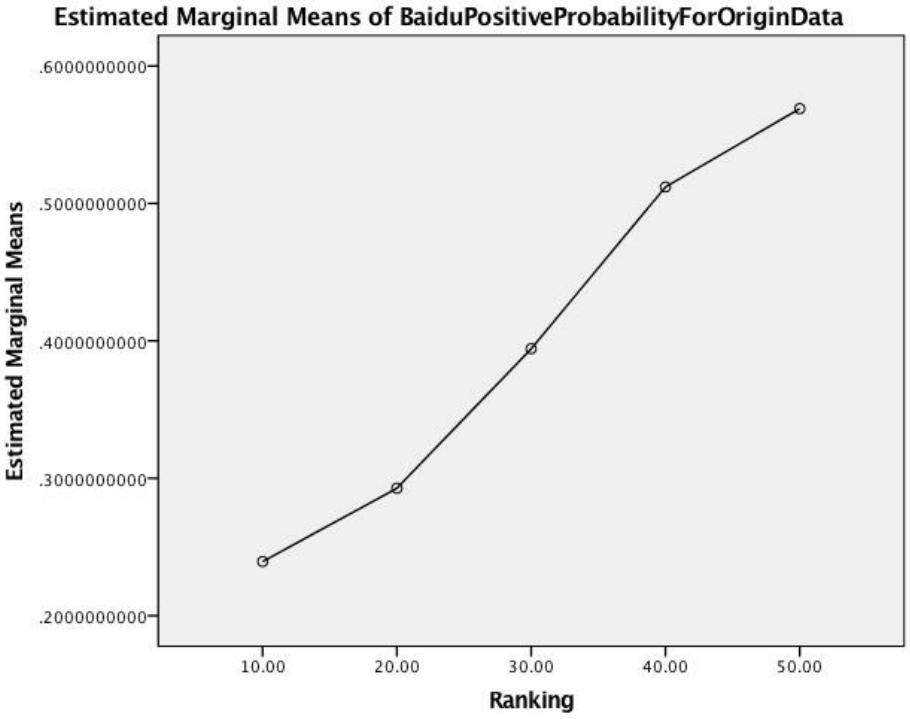
# 2 Baidu Chinese sentiment analysis

## 2.1 Baidu Chinese sentiment analysis positive probability histogram



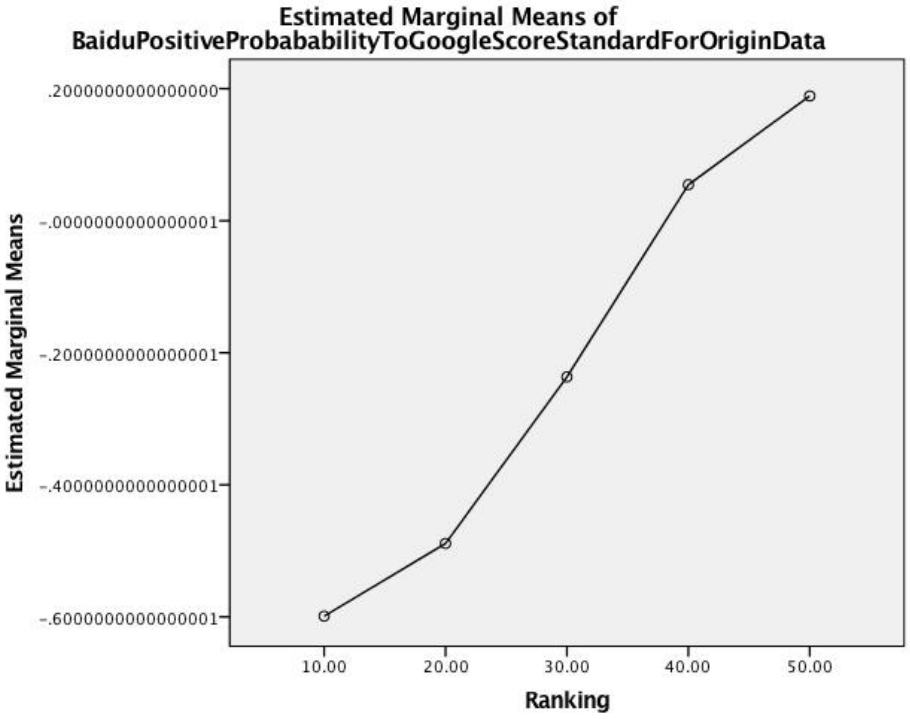

## 2.2 Baidu Chinese sentiment analysis postitive probability compare with different ranking(origin data)

| Ranking | Mean | Valid N | std.deviation | Total N | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Ranking 10 | 0.239365965000 | 8525 | 0.2219412270000 | 8572 | 0.000106 | 1.000000 |
| Ranking 20 | 0.292751426000 | 13141 | 0.2357115580000 | 13226 | 0.000162 | 1.000000 |
| Ranking 30 | 0.394234 | 18821 | 0.273685 | 18974 | 0.000214 | 1.000000 |
| Ranking 40 | 0.511990 | 8717 | 0.300618 | 8790 | 0.001050 | 1.000000 |
| Ranking 50 | 0.568988 | 4271 | 0.312815 | 4307 | 0.000536 | 1.000000 |




## 2.3 Baidu Chinese sentiment analysis postitive probability tranform to Google Score standard compare with different ranking (origin data)

| Ranking | Mean | Valid N | std.deviation | Total N | Minimum | Maximum | Variance |
|---|---|---|---|---|---|---|---|
| Ranking 10 | -0.598875 | 8525 | 0.557595 | | -0.999894 | 1.000000 | 0.310912 |
| Ranking 20 | -0.488772 | 13141 | 0.617021 | | -0.999838 | 1.000000 | 0.380715 |
| Ranking 30 | -0.236524 | 18821 | 0.728420 | | -0.999786 | 1.000000 | 0.530596 |
| Ranking 40 | 0.054493 | 8717 | 0.773410 | | -0.998950 | 1.000000 | 0.598164 |
| Ranking 50 | 0.188983 | 4271 | 0.774245 | | -0.999464 | 1.000000 | 0.599456 |
| Total | -0.274854 | 53475 | 0.733884 | | -0.999894 | 1.000000 | 0.538586 |




### 2.3.1 Error Rate

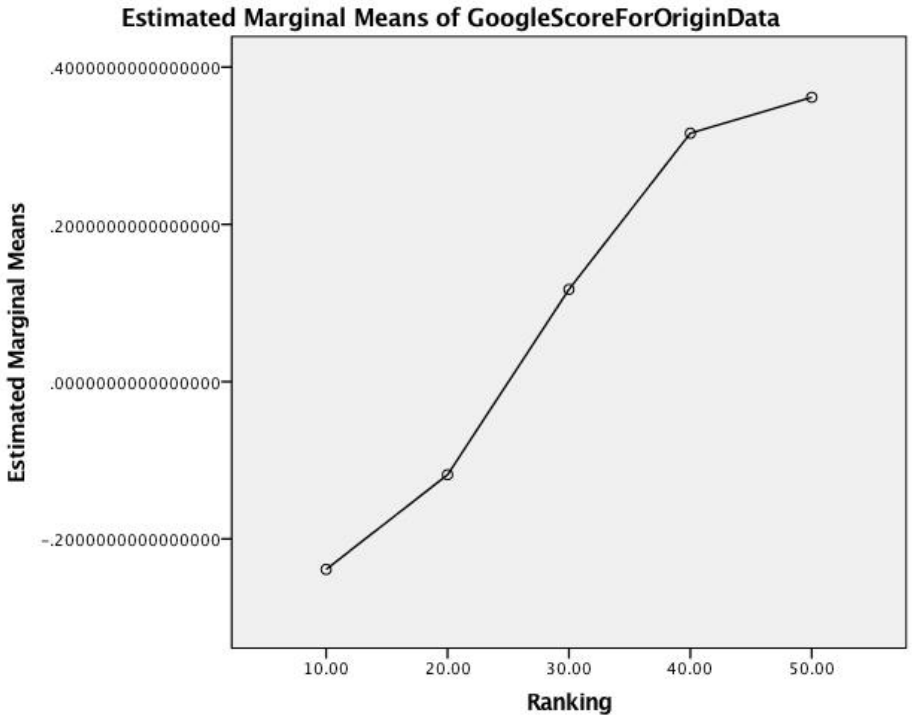
| Ranking | Error Rate |
|---|---|
| Ranking 10 | 0.0054829678 |
| Ranking 20 | 0.0064267352 |
| Ranking 30 | 0.0080636661 |
| Ranking 40 | 0.0083048919 |
| Ranking 50 | 0.0083584862 |

Total Error Rate: 0.0073140396

# 3 Google Chinese sentiment analysis

## 3.1 Google Chinese sentiment analysis scores compare with different ranking (origin data)

| Ranking | Mean | Valid N | std.deviation | Total N | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Ranking 10 | -0.238742 | 8567 | 0.445384 | 8572 | -0.900000 | 0.900000 |
| Ranking 20 | -0.118380 | 13210 | 0.448064 | 13226 | -0.900000 | 0.900000 |
| Ranking 30 | 0.117291 | 18940 | 0.462095 | 18974 | -0.900000 | 0.900000 |
| Ranking 40 | 0.315915 | 8778 | 0.458128 | 8790 | -0.900000 | 0.900000 |
| Ranking 50 | 0.361626 | 4305 | 0.441309 | 4307 | -0.900000 | 0.900000 |




### 3.1.1 Error Rate

| Ranking | Error Rate |
|---|---|
| Ranking 10 | 0.0005832944 |
| Ranking 20 | 0.0012097384 |
| Ranking 30 | 0.0017919258 |
| Ranking 40 | 0.0013651877 |
| Ranking 50 | 0.0004643603 |

Total Error Rate: 0.0012808851

# 4 Baidu Chinese sentiment analysis VS Google Chinese sentiment analysis

## 4.1 Mean Value Correlation

- Pearson Correlation 0.991
- sig. 0.001
- N 5
- Conclusion Baidu Chinese sentiment analysis and Google Chinese sentiment analysis have higher liner relationship.

## 4.2 Error Rate

| Baidu Chinese sentiment analysis Total Error Rate | Google Chinese sentiment analysis Total Er |
|---|---|
| 0.0073140396 | 0.00 |

- conclusion
  - Baidu sentiment analysis error rate high than Google sentiment analysis error rate

## 4.3 Tendency

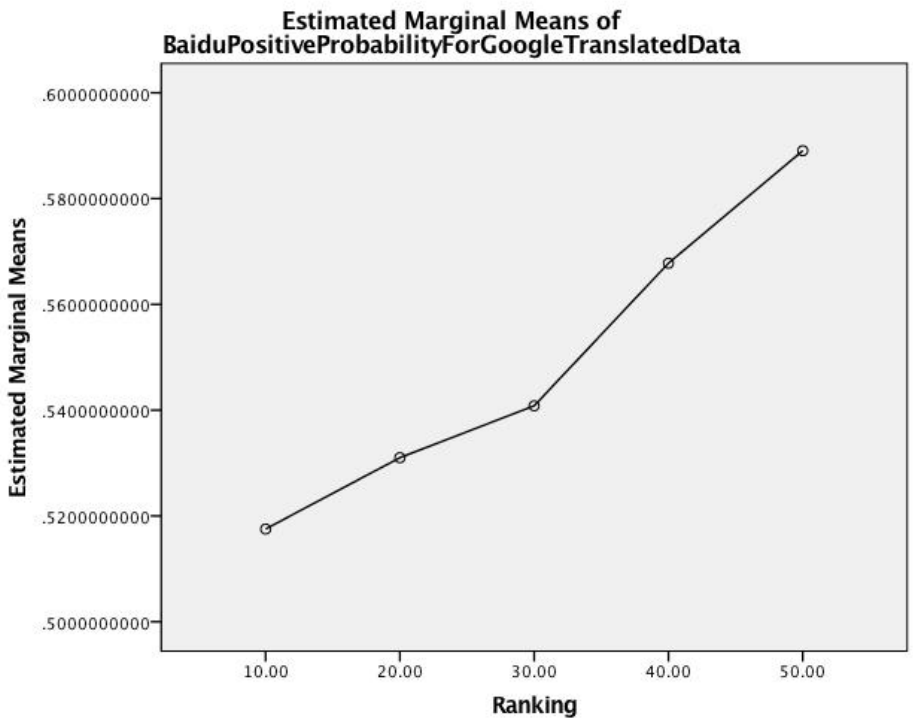
sentiment analysis results given by both Baidu and Google are valid because when the ranking group ID increases from 10 to 50, the sentiment analysis score also strictly increases accordingly.

# 5 Baidu English sentiment analysis VS Google English sentiment analysis

## 5.1 Base on Google translated data

### 5.1.1 Baidu English sentiment analysis postitive probability compare with different ranking(based on Google translated data)

| Ranking | Mean | Valid N | Std.deviation | Total N | Minimum | Maximum | Variance |
|---|---|---|---|---|---|---|---|
| Ranking 10 | 0.517526 | 7968 | 0.134711 | | 0.005045 | 1.000000 | 0.018147 |
| Ranking 20 | 0.531020 | 12225 | 0.141214 | | 0.037275 | 1.000000 | 0.019941 |
| Ranking 30 | 0.540824 | 17457 | 0.137174 | | 0.014443 | 1.000000 | 0.018817 |
| Ranking 40 | 0.567782 | 8163 | 0.144971 | | 0.051860 | 1.000000 | 0.021016 |
| Ranking 50 | 0.589054 | 4006 | 0.150737 | | 0.086614 | 1.000000 | 0.022722 |



### 5.1.2 Baidu English sentiment analysis positive tranform to Google Score Standard (based on Google translated data)