

1. 与方翔的数据进行了对比，找出了127个有误的数据，并分析了我这边计算错误的原因

Standard Bank	STANDARD BANK	STANDARD BANK GROUP LTD
SVB Financial Group	SVB FINANCIAL GROUP	JB FINANCIAL GROUP CO LTD
SVB Financial Group	SVB FINANCIAL GROUP	JB FINANCIAL GROUP CO LTD
Qiagen Corporation	QIAGEN NV	AMGEN INC
Newpage corporation	NEWPAGE CORP	NEWAG SA
Pfizer Inc	PFIZER LTD	PFIZER INC
Premier Inc	PREMIER LTD	PREMIER INC
Premier	PREMIER LTD	PREMIER INC
SCANA Corp	SCANA ASA	SCANA CORP
Earthstone Energy Inc	EARTHSTONE ENERGY INC	EARTHSTONE ENERGY INC -OLD
Pfizer Inc	PFIZER LTD	PFIZER INC
American Airlines Inc	AMERICAN AIRLINES INC	AMERICAN AIRLINES GROUP INC
BRP Inc	BRP INC	BRP GROUP INC

数据有误的原因：

1. 在第三步中，删除了GROUP后缀的影响，但还有Group，group的影响没有删除，重新清理。
 2. 数据匹配不够仔细，在对清理后的数据匹配之后，也应该对缩写公司的类型进行判断，所以重新再细分了匹配细则，对公司后缀和公司地区进行比较，会让数据更准确
2. 在以上基础上对代码进行了改进。
3. 由于数据匹配方法不同，导致两人的部分匹配数据不具有可比性：
- 我是以公司名称和缩写字符串的相似度进行匹配
- 方翔以公司名称中包含缩写字符串进行匹配
- 因此匹配结果有些差异很大，对两人来说都不存在可参考性

关于此部分解决方法：

以我匹配度最高的名称与他那边给出的Final Result与缩写再进行相似度比较，保留相似度高的，而后再对最终结果进行进一步判断。