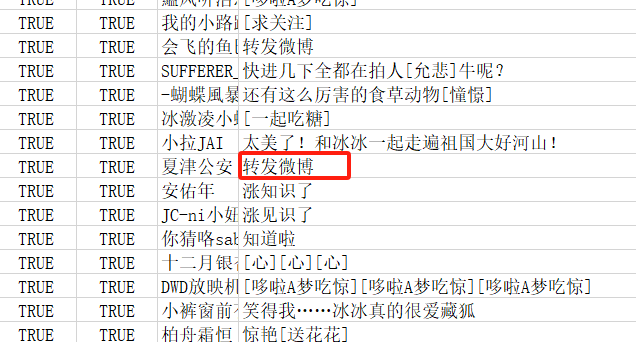
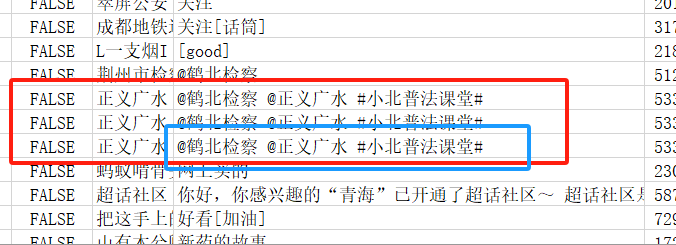
用评论内容进行分析。

数据处理：

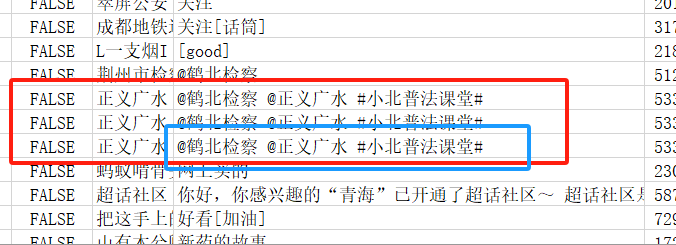
1.评论内容仅为“转发微博”视为无效数据，整条删掉。



2.重复数据去重



[3.对于@用户名的评论需要做处理](mailto:3.对于@用户名的评论需要做处理)，把评论里的@用户删掉



数据分析：

首先对所有数据进行LDA主题识别，得出x个主题(用pyLDAvis可视化后要求每个气泡大致均匀，能找到最佳主题数)。

每个主题需要能归纳出一个评价角度（例如生物量、水源等，这个可以把识别出的主题的关键词给我看，我自己归纳）然后给每条评论打上主题标签。

接着将数据按照不同公园分类，再将一个公园的评论数据按照不同主题分类做情感分析，（例如识别出四个主题，对每一个公园都要从这四个角度进行情感分类，一共五个公园，就要做4x5=20次情感分类），这样就能达到对五个公园从不同角度评价。

得出情感分类结果之后，按照什么指标或者计算方法，得到一个评分。每个公园的每个角度都有一个评分