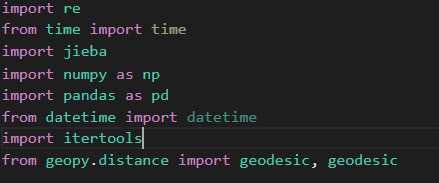
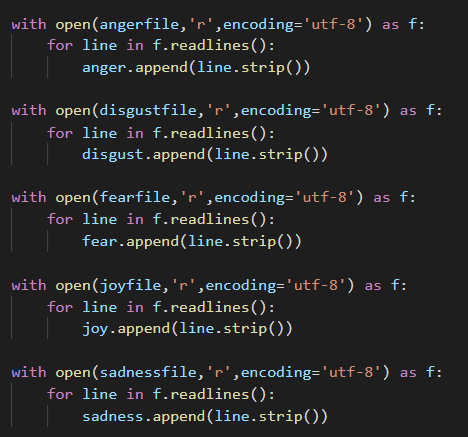
**初始化：**





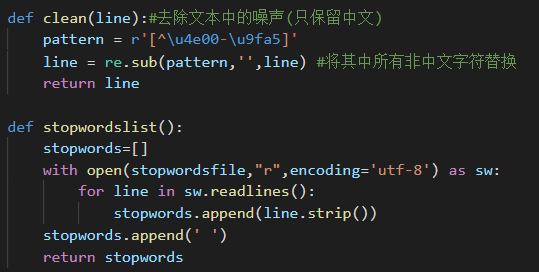
**创建情绪词典：**



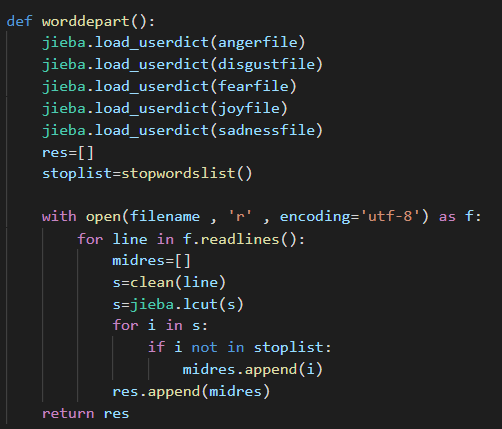


1. 实现一个函数，对微博数据进行清洗，去除噪声（如url等），过滤停用词。注意分词的时候应该将情绪词典加入Jieba的自定义词典，以提高这些情绪词的识别能力。

**创建停用词列表，并且事先实现去除噪声的功能，这步只考虑分词，过去除噪声只保留中文**



实现分词的函数：

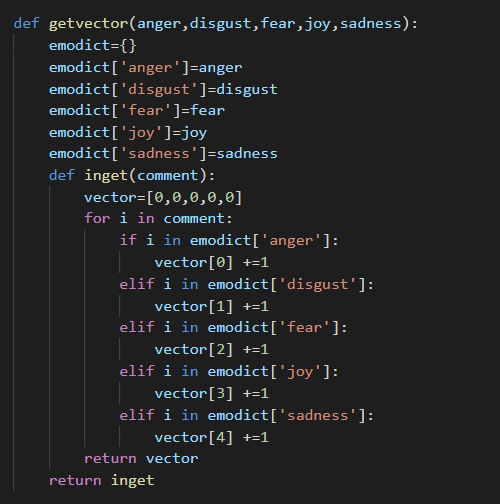


运行结果：

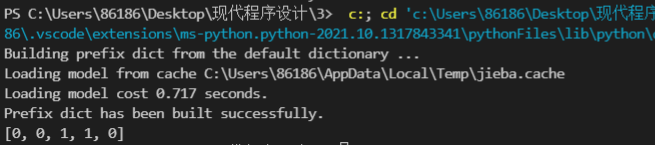


1. 实现两个函数，实现一条微博的情绪分析，返其情绪向量或情绪值。目前有两种方法，一是认为一条微博的情绪是混合的，即一共有n个情绪词，如果joy有n1个，则joy的比例是n1/n；二是认为一条微博的情绪是唯一的，即n个情绪词里，anger的情绪词最多，则该微博的情绪应该为angry。注意，这里要求用闭包实现，尤其是要利用闭包实现一次加载情绪词典且局部变量持久化的特点。同时，也要注意考虑一些特别的情况，如无情绪词出现，不同情绪的情绪词出现数目一样等，并予以处理（如定义为无情绪，便于在后面的分析中去除）。

**方法一：情绪向量**

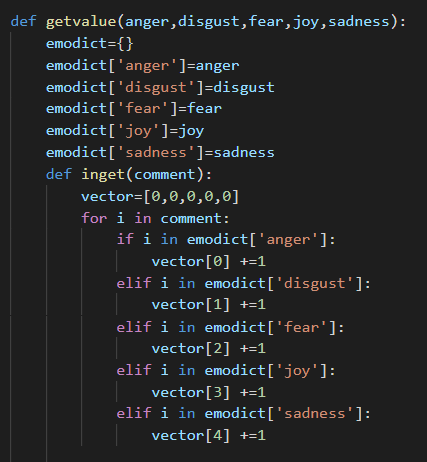


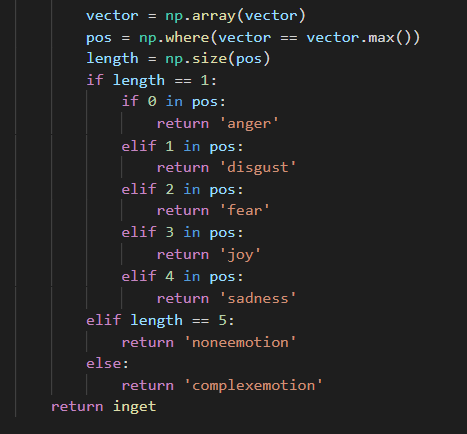
**运行结果：**



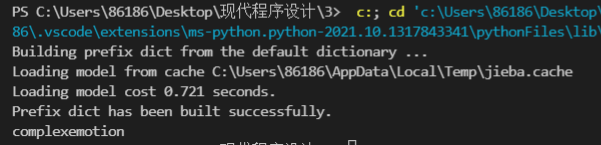
**方法二：唯一值**

**是认为一条微博的情绪是唯一的，即n个情绪词里，anger的情绪词最多，则该微博的情绪应该为angry，如果没有情绪词则为nonemotion，如果最大情绪词数目有多个则为complexemotion**



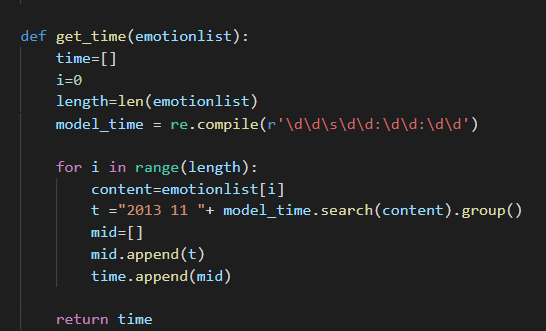


**运行结果：**



1. 微博中包含时间，可以讨论不同时间情绪比例的变化趋势，实现一个函数，可以通过参数来控制并返回对应情绪的时间模式，如joy的小时模式，sadness的周模式等。

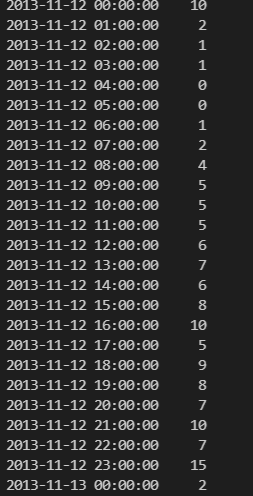
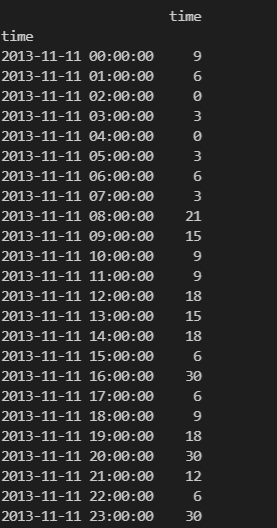
**为了实现这个功能，首先要将每条微分类，一共有七种分别是anger，disgust，fear，joy，sadness，noneemotion，complexemotion。用列表储存微博的原文本。然后利用正则表达式将某个情绪的时间提取出来，并利用列表储存**



**然后利用pandas库来统计某个情绪在不同时间段的频数，model为不同的模式，0为每半小时，1为每小时，2为每天，emotype为不同情绪类型，0为anger，1为disgust，2为fear，3为joy，4为sadness**

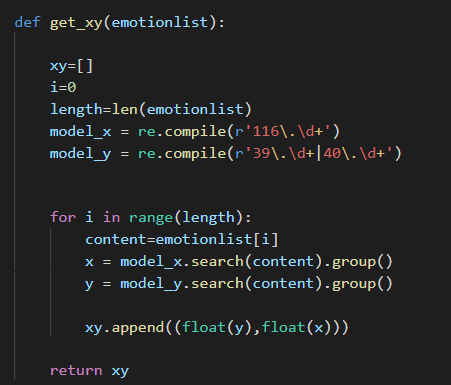


**运行结果：**



1. 微博中包含空间，可以讨论情绪的空间分布，实现一个函数，可以通过参数来控制并返回对应情绪的空间分布，即围绕某个中心点，随着半径增加该情绪所占比例的变化，中心点可默认值可以是城市的中心位置。

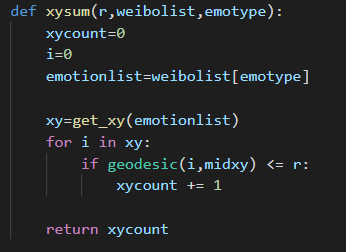
**首先获取每条评论的坐标，储存在列表中**



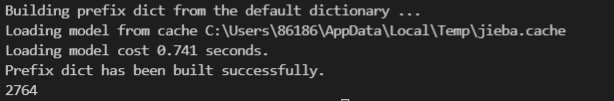
**定义北京市中心的坐标：**



**然后计算每条评论的坐标距离中心的位置，小于半径r的进行统计，emotype为不同情绪类型，0为anger，1为disgust，2为fear，3为joy，4为sadness**



**运行结果：**



**主函数：**

