给纽约时报编辑的信：

我们发现了给定的数据集中有几个异常值，有的字符串只有四个字母，也有的单词的报告数量与前后日期对比差距过多，我们将以上情况视为异常值，首先对其进行了剔除。

为了对每天的报告结果的数量进行预测，我们选择了使用拟合的方法来进行预测。通过该方法，我们最终得到一个拟合公式，通过公式，我们代入要预测的日期，即可得到想要的预测结果。我们设定了一个预测区间来表示我们的预测结果。通过观察报告数量的变化趋势我们不难发现，前面很大一部分的数据会对我们的预测产生一定程度的干扰。为了得到预测更为精准的公式，因此我们在拟合时去除了早期的一部分数据。

另外，我们为单词制定了单词的词性、单词使用频率、元音字母个数、首字母等属性，通过对这些属性与在困难模式下播放的分数的百分比进行可视化分析，我们可以得出哪些属性对在困难模式下播放的分数的百分比是有明显的影响的。通过对结果的分析，我们最终选择了单词的使用频率、元音字母的个数、字母频率以及单词所拥有的字母的种类数作为特征来进行下面的分析。其他的属性由于对结果的影响不大，我们选择将其舍弃。

基于我们自己建立的模型，我们对EERIE这个单词的难度等级进行了预测，最终得出它属于中等难度，我们计算出它大概平均四次或者五次可以被猜出。通过对已知数据集进行划分并检验，我们发现我们的模型拥有80.4%的准确率，我们对这个单词的预测结果是否准确只有实践结果来检验。

除此之外，我们还发现了一些该数据集的有趣的属性。我们使用在困难模式下播放的分数的百分比除以总报告数量，得到了一个比值HMR(Hard Mode Rate)。我们将HMR随时间的变化可视化地表示了出来，最终发现困难模式得分占比在缓慢提高，最终稳定在了一个稳定的水平，我们推测这可能与游戏玩法被开发或者游戏玩家的水平提高有很大的关系。

在近几天我们每天早上都会打开Wordle的网站一起来玩这个有趣的游戏，在游戏的过程中，我们也总结出了一些浅显的玩法，对此我们感到很开心。尽管有些时候我们并不能猜出来答案，但是它确实使我们认识了一些比较有意思的单词！