**Problem C: Predicting Wordle Results**

**问题C：预测单词结果**

**背景**

Wordle是《纽约时报》目前每天提供的流行谜题。玩家尝试解决

通过在六次或更少的尝试中猜测一个五个字母的单词来拼图，每次猜测都会收到反馈。

对于此版本，每个猜测必须是英语中的实际单词。无法识别的猜测

因为比赛的文字是不允许的。Wordle的受欢迎程度和版本继续增长

该游戏现在有 60 多种语言版本。

《纽约时报》网站对Wordle的指示指出，瓷砖的颜色将发生变化。

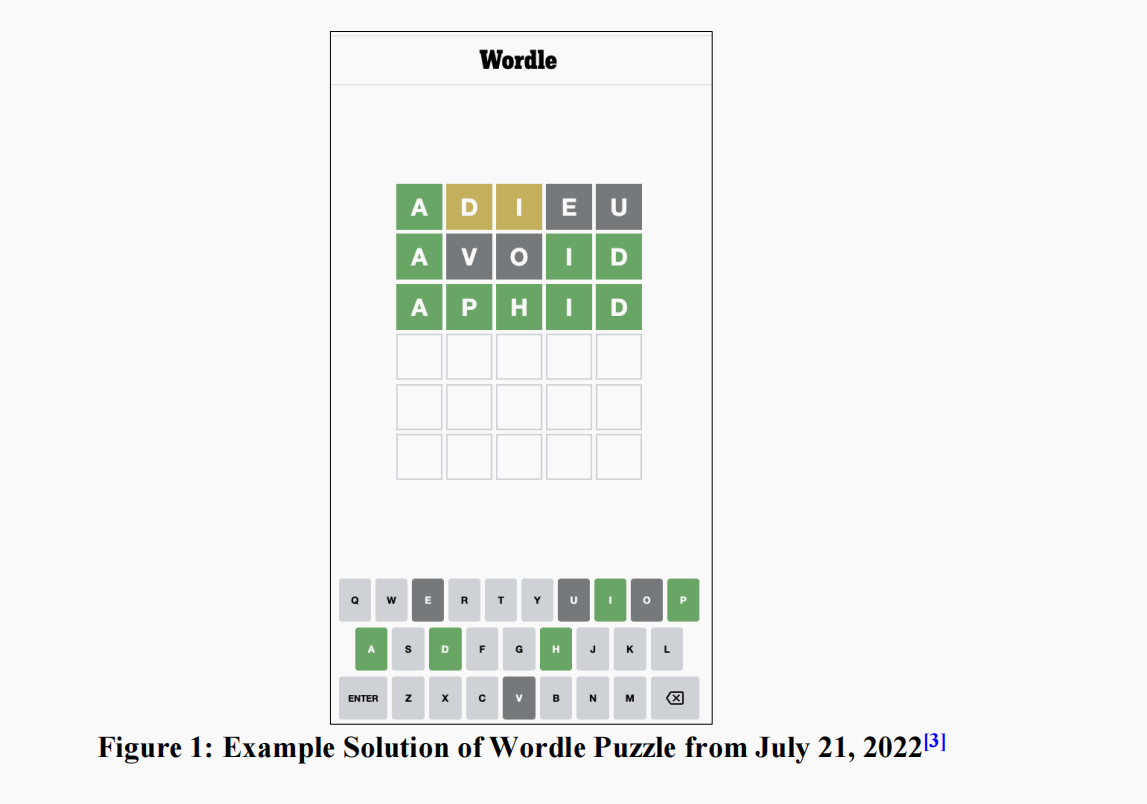
在你提交你的话之后。黄色磁贴表示该磁贴中的字母在单词中，但它位于

位置错误。绿色磁贴表示该磁贴中的字母位于单词中，并且位于

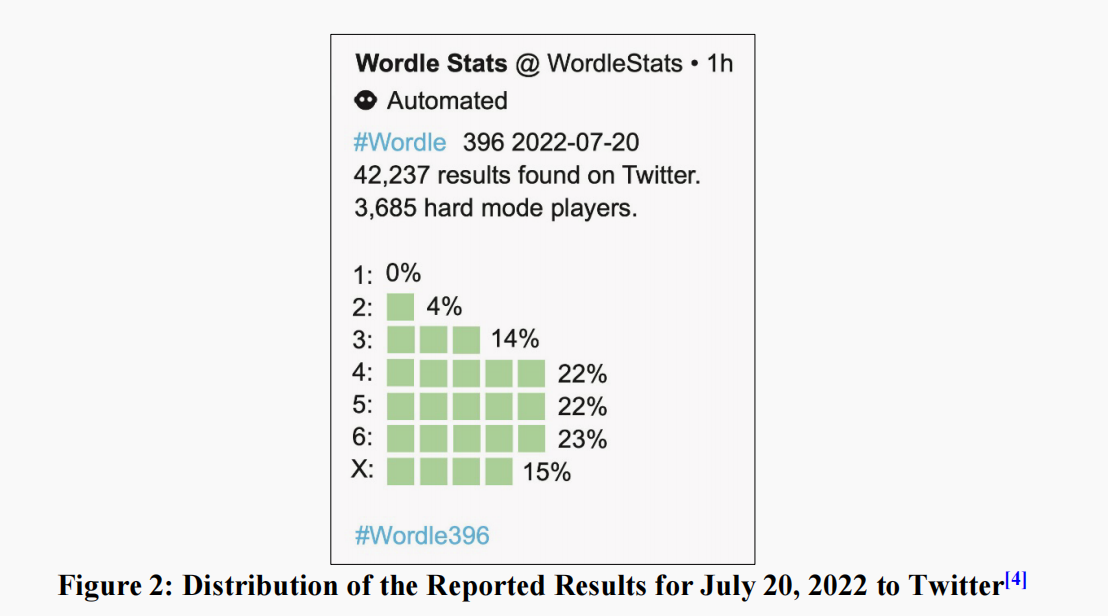
正确的位置。灰色磁贴表示该磁贴中的字母根本不包含在单词中

（见附件2）

图1是一个示例解决方案，在三次尝试中都找到了正确的结果。



玩家可以在常规模式或“硬模式”下玩。Wordle的硬模式使得游戏更加困难，因为它要求一旦玩家在一个单词中找到了一个正确的字母（平铺是黄色或绿色的），这些字母就必须在随后的猜测中使用。图1中的示例是在硬模式下播放的。许多（但不是所有的）用户都在推特上报告了他们的分数。对于这个问题，MCM已经生成了2022年1月7日至2022年12月31日的每日结果文件（见附件1）。这个文件包括日期、比赛号码，一天的单词，报告分数的人数，球员的硬模式，和猜测的比例在一次尝试，两次尝试，三次尝试，四次尝试，五次，六次尝试，或无法解决谜题（由X）。例如，在图2中，2022年7月20日的单词是“陈腐的”，结果是通过挖掘Twitter得到的。虽然图2中的百分比总和为100%，但在某些情况下，由于四舍五入，这可能不是真的。



**要求**

纽约时报要求您对此文件的结果进行分析，以回答几个问题。

•报告的结果的数量每天都在变化。开发一个模型来解释这种变化，并使用您的模型为2023年3月1日报告的结果数量创建一个预测区间。单词的任何属性是否会影响在困难模式下播放的分数的百分比？如果是，如何处理？如果不是，为什么不呢？

•对于未来日期的给定未来解决方案词，开发一个模型，允许您预测报告结果的分布。换句话说，来预测未来一个日期的（1、2、3、4、5、6、X）的相关百分比。你的模型和预测有哪些不确定性？举一个你在2023年3月1日预测怪异这个词的具体例子。你对你的模型的预测有多有信心？

·开发和总结一个模型，按难度分类解决方案词。识别与每个分类关联的给定单词的属性。使用你的模型，怪诞（ERRIE）这个词有多难？讨论你的分类模型的准确性。

•列出并描述了这个数据集的其他一些有趣的特性。

最后，在一封给《纽约时报》拼图编辑的一到两页的信中总结你的结果。

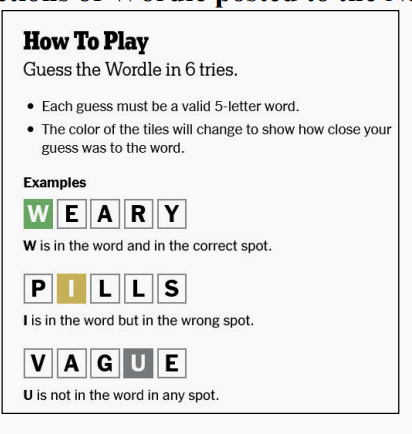
您的PDF解决方案总共不超过25页，应包括：一页纸的摘要表。•产品目录表。•你的完整的解决方案。•一到两页的信。•参考资料列表。

**附件**

1。数据文件。问题C数据Wordle.xlsx附加的数据文件包含了针对此问题应该使用的唯一数据。此问题所需的所有信息都在问题语句和数据文件中给出。你不需要访问纽约时报的网站或推特网站。在这些网站上没有找到其他的信息。

数据文件输入描述日期：给定字谜的mm-dd-yyyy（月-日）格式的日期。比赛编号：Wordle谜题索引，2022年1月7日202开始。Word：解决方案字玩家正试图猜测相关的日期和比赛号码。报告结果的数量：当天在推特上记录的总分。硬模式下的数字：当天在推特上记录的硬模式下的分数的数量。1尝试：玩家猜谜题的百分比。2次尝试：在两次猜测中，玩家解决谜题的百分比。3次尝试：玩家在三次猜测中解决谜题的百分比。4次尝试：玩家四次猜出谜题的百分比。5次尝试：玩家解决谜题的百分比。6次尝试：玩家6次猜谜的百分比。7次或更多尝试(X)：6次或更少尝试无法解决难题的玩家百分比。注意：由于四舍五入，这个百分比可能并不总是等于100%。

1. Wordle的方向发布在了纽约时报的网站上。



1. 纽约时报：美国纽约纽约市的日报，在印刷和网上出版。推特：一个社交网站，允许用户播放不超过280个字符的短片（从最初的140个字符增加到现在）。解决（字谜）：按正确的顺序输入正确的字母，形成当天的字谜。