2023 MCM
Problem C: Predicting Wordle Results

Image: nytco.com[1]

背景：

Wordle是目前由《纽约时报》每日提供的一款流行的谜题。玩家尝试在六次或更少的猜测中猜出一个由五个字母组成的单词，并且每次猜测都会得到反馈。在这个版本中，每次猜测都必须是英语中的一个实际单词。不被比赛认可为单词的猜测是不被允许的。Wordle继续在人们中间流行，游戏的版本现在已经有60多种语言可用。

《纽约时报》网站的Wordle说明称，在提交单词后，瓷砖的颜色会发生变化。黄色的瓷砖表示该瓷砖中的字母是单词中的字母，但位置错误。绿色的瓷砖表示该瓷砖中的字母是单词中的字母，且位置正确。灰色的瓷砖表示该瓷砖中的字母在单词中根本不存在（见附件2）。图1是一个例子，其中在三次尝试中找到了正确答案。

Figure 1: Example Solution of Wordle Puzzle from July 21, 2022[3]

玩家可以在普通模式或“困难模式”下进行游戏。Wordle的困难模式通过要求一旦玩家在单词中找到了一个正确的字母（瓷砖是黄色或绿色），那些字母必须在随后的猜测中使用，从而使游戏变得更加困难。图1的示例是在困难模式下进行的。

许多用户（但不是所有用户）在Twitter上报告他们的分数。对于这个问题，MCM已经生成了一个文件，涵盖了2022年1月7日到12月31日的每天的结果（见附件1）。该文件包括日期、比赛编号、当天报告分数的人数、困难模式下的玩家人数，以及在一次、两次、三次、四次、五次、六次中猜出单词或无法解决谜题的玩家所占的百分比。例如，在图2中，2022年7月20日的单词是“TRITE”，结果是通过Twitter挖掘获得的。尽管图2中的百分比总和为100%，但在某些情况下，由于四舍五入，这可能不是真实的

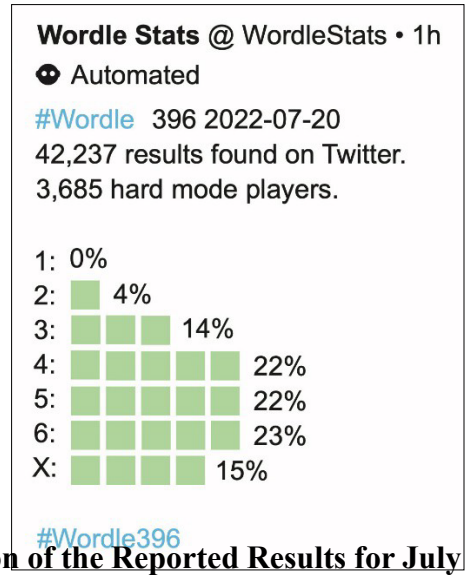


Figure 2: Distribution of the Reported Results for July 20, 2022 to Twitter[4]

要求：

纽约时报要求您对此文件中的结果进行分析，回答几个问题。

• 报告的结果数量每天都有所变化。开发一个模型来解释这种变化，并使用您的模型创建一个用于预测2023年3月1日报告结果数量的预测区间。单词的哪些属性会影响在困难模式下玩的得分所占百分比？如果有，是如何影响的？如果没有，为什么没有？

• 对于未来日期的给定解决方案单词，开发一个模型，使您能够预测报告结果的分布。换句话说，预测未来日期的（1、2、3、4、5、6、X）所占的相关百分比。您的模型和预测有哪些不确定性？给出一个关于2023年3月1日单词EERIE的具体预测示例。您对模型的预测有多大信心？

解释：

题目要求针对Wordle游戏的数据进行分析，并回答两个问题。第一个问题要求我们开发一个模型来解释每天报告结果数量的变化，并使用模型预测2023年3月1日的报告结果数量的预测区间。其次，需要分析哪些单词属性会影响在困难模式下玩的得分所占百分比。第二个问题要求我们开发一个模型来预测未来日期的单词的报告结果分布，并分析预测的不确定性。最后，需要给出一个具体的预测示例，并评估模型预测的准确性。

• 开发并总结一个模型，以分类解决方案单词的难度。确定与每个分类相关的给定单词的属性。使用您的模型，单词EERIE有多难？讨论您分类模型的准确性。

• 列出并描述数据集中的一些其他有趣特征。

最后，用一到两页的信件向《纽约时报》谜题编辑总结您的结果。

- Develop and summarize a model to classify solution words by difficulty. Identify the attributes of a given word that are associated with each classification. Using your model, how difficult is the word EERIE? Discuss the accuracy of your classification model.

- List and describe some other interesting features of this data set.

Finally, summarize your results in a one- to two-page letter to the Puzzle Editor of the New York Times.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.
- One- to two-page letter.
- Reference List.

Note: The MCM Contest has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, Report, Reference List, and any Appendices). You must cite the sources for your ideas, images, and any other materials used in your report.

Attachments

1. Data File. Problem C Data Wordle.xlsx

THE ATTACHED DATA FILE CONTAINS THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM. All information needed for this problem is given in the problem statement and the data file. You do not need to visit the New York Times website nor Twitter website. There is no additional information to be found on these sites.

Data File Entry Descriptions

Date: The date in mm-dd-yyyy (month-day-year) format of a given Wordle puzzle.

Contest number: An index of the Wordle puzzles, beginning with 202 on January 7, 2022.

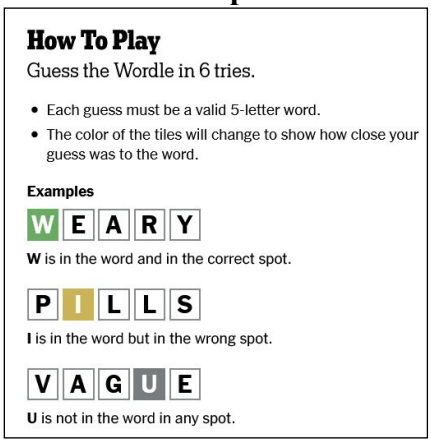
Word: The solution word players are trying to guess on the associated date and contest number.

Number of reported results: The total number scores that were recorded on Twitter that day.

Number in hard mode: The number of scores on Hard mode recorded on Twitter that day.

1 try: The percentage of players solving the puzzle in one guess.
2 tries: The percentage of players solving the puzzle in two guesses.
3 tries: The percentage of players solving the puzzle in three guesses.
4 tries: The percentage of players solving the puzzle in four guesses.
5 tries: The percentage of players solving the puzzle in five guesses.
6 tries: The percentage of players solving the puzzle in six guesses.
7 or more tries (X): The percentage of players that could not solve the puzzle in six or fewer tries. Note: the percentages may not always sum to 100% due to rounding.

2. Directions of Wordle posted to the New York Times website.[2]



Glossary

New York Times: A daily newspaper based in New York City, New York, USA published in print and online.

Twitter: A social networking site that allows users to broadcast short posts of no more than 280 characters (increased from initial 140 characters).

Solve (the Wordle puzzle): Enter the correct letters in the correct order to form the Wordle word of the day.

References

Note: We provide the following citations to support the Problem Statement. We have pulled the important ideas from these resources. There is no additional information on these websites needed to solve this MCM problem. Access to the New York Times or Twitter website is not required to solve this problem.

[1] Wordle logo from The New York Times website. Accessed on December 13, 2022 at https://nytco-assets.nytimes.com/2022/08/cropped-Screen-Shot-2022-08-24-at-8.49.39-AM.png.

[2] "Wordle-The New York Times." The New York Times, 2022. Accessed December 13, 2022 at https://www.nytimes.com/games/wordle/index.html.

[3] "Wordle-The New York Times." The New York Times, July 21, 2022.

[4] "Wordle Stats." Twitter, July 20, 2022.