

2023 MCM 问题 C: 预测 Wordle 结果



Image: nytco.com¹¹ 图像: nytco.com¹¹

Background 背景

Wordle is a popular puzzle currently offered daily by the New York Times. Players try to solve the puzzle by guessing a five-letter word in six tries or less, receiving feedback with every guess. For this version, each guess must be an actual word in English. Guesses that are not recognized as words by the contest are not allowed. Wordle continues to grow in popularity and versions of the game are now available in over 60 languages.

Wordle 是纽约时报目前每天提供的一个受欢迎的谜题。玩家尝试在六次尝试内猜出一个五个字母的单词，并在每次猜测后获得反馈。在这个版本中，每次猜测必须是一个实际的英语单词。比赛不允许不被认可为单词的猜测。Wordle 的受欢迎程度持续上升，游戏的版本现在已提供超过 60 种语言。

The New York Times website directions for Wordle state that the color of the tiles will change after you submit your word. A yellow tile indicates the letter in that tile is in the word, but it is in the wrong location. A green tile indicates that the letter in that tile is in the word and is in the correct location. A gray tile indicates that the letter in that tile is not included in the word at all (see Attachment 2)^[2]. Figure 1 is an example solution where the correct result was found in three tries.

《纽约时报》网站关于 Wordle 的说明指出，提交单词后，瓷砖的颜色会发生变化。黄色瓷砖表示该瓷砖中的字母在单词中，但位置错误。绿色瓷砖表示该瓷砖中的字母在单词中且位置正确。灰色瓷砖表示该瓷砖中的字母根本不包含在单词中（见附件 2）^[2]。图 1 是一个示例解决方案，其中在三次尝试中找到了正确的结果。

Wordle



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

图 1: 2022 年 7 月 21 日 Wordle 谜题的示例解答 ^[3]

Players can play in regular mode or “Hard Mode.” Wordle’s Hard Mode makes the game more difficult by requiring that once a player has found a correct letter in a word (the tile is yellow or green), those letters must be used in subsequent guesses. The example in Figure 1 was played in Hard Mode.

玩家可以选择常规模式或“困难模式”。Wordle 的困难模式通过要求一旦玩家在单词中找到了正确的字母（该字母为黄色或绿色），这些字母必须在后续的猜测中使用，从而使游戏变得更加困难。图 1 中的示例是在困难模式下进行的。

Many (but not all) users report their scores on Twitter. For this problem, MCM has generated a file of daily results for January 7, 2022 through December 31, 2022 (see Attachment 1). This file includes the date, contest number, word of the day, the number of people reporting scores that day, the number of players on hard mode, and the percentage that guessed the word in one try, two tries, three tries, four tries, five tries, six tries, or could not solve the puzzle (indicated by X). For example, in Figure 2 the word on July 20, 2022 was “TRITE” and the results were obtained by mining Twitter. Although the percentages in Figure 2 sum to 100%, in some cases this may not be true due to rounding.

许多（但不是全部）用户在 Twitter 上报告他们的分数。对于这个问题，MCM 生成了一个从 2022 年 1 月 7 日到 2022 年 12 月 31 日的每日结果文件（见附件 1）。该文件包括日期、比赛编号、每日单词、当天报告分数的人数、困难模式下的玩家人数，以及一次尝试、两次尝试、三次尝试、四次尝试、五次尝试、六次尝试或无法解答谜题（用 X 表示）的人所占的百分比。例如，在图 2 中，2022 年 7 月 20 日的单词是“TRITE”，结果是通过挖掘 Twitter 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，这可能不成立。

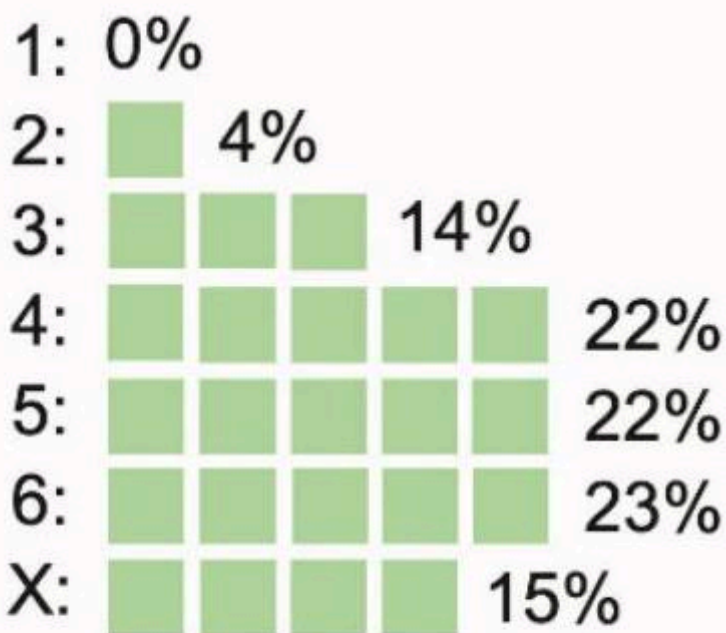
Wordle Stats @ WordleStats • 1h

🤖 Automated

#Wordle 396 2022-07-20

42,237 results found on Twitter.

3,685 hard mode players.



#Wordle396

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

图 2: 2022 年 7 月 20 日到 Twitter 的报告结果分布 ^[4]

Requirement 需求

You have been asked by the New York Times to do an analysis of the results in this file to answer several questions.

你被《纽约时报》要求对这个文件中的结果进行分析，以回答几个问题。

The number of reported results vary daily. Develop a model to explain this variation and use your model to create a prediction interval for the number of reported results on March 1, 2023. Do any attributes of the word

affect the percentage of scores reported that were played in Hard Mode? If so, how? If not, why not?

报告结果的数量每天都在变化。开发一个模型来解释这种变化，并使用你的模型为 2023 年 3 月 1 日的报告结果数量创建一个预测区间。单词的任何属性是否影响以困难模式进行的得分报告的百分比？如果是，怎么影响？如果不是，为什么不是？

For a given future solution word on a future date, develop a model that allows you to predict the distribution of the reported results. In other words, to predict the associated percentages of (1, 2, 3, 4, 5, 6, X) for a future date. What uncertainties are associated with your model and predictions? Give a specific example of your prediction for the word EERIE on March 1, 2023. How confident are you in your model's prediction?

对于未来某个日期的给定解决方案词，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) 相关百分比。您的模型和预测有哪些不确定性？请给出 2023 年 3 月 1 日对单词 EERIE 的预测的具体示例。您对模型的预测有多大的信心？

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.

开发并总结一个模型，以根据难度对解决方案词进行分类。识别与每个分类相关的给定单词的属性。使用你的模型，单词 EERIE 的难度如何？讨论你的分类模型的准确性。

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

列出并描述该数据集的一些其他有趣特征。

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

最后，将您的结果总结成一封一到两页的信件，寄给《纽约时报》的谜题编辑。

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

您的 PDF 解决方案总页数不超过 25 页，应包括：

One-page Summary Sheet. 一页摘要表。

Table of Contents. 目录。

Your complete solution. 您的完整解决方案。

One- to two-page letter.

一到两页的信。

Reference List. 参考文献列表。

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.

注意：MCM 竞赛有 25 页的限制。您提交的所有内容都计入 25 页的限制（摘要表、目录、报告、参考文献列表和任何附录）。您必须引用您在报告中使用的想法、图像和任何其他材料的来源。

Attachments 附件

1. Data File. Problem C Data Wordle.xlsx

1. 数据文件。问题 C 数据 Wordle.xlsx

THE ATTACHED DATA FILE CONTAINS THE ONLY DATA YOU SHOULD USE FOR THIS

附加的数据文件包含您应该用于此的唯一数据

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.

问题。解决此问题所需的所有信息均在问题陈述和数据文件中提供。您无需访问纽约时报网站或推特网站。这些网站上没有额外的信息可供查找。

Data File Entry Descriptions

数据文件条目描述

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

日期：给定 Wordle 谜题的日期，格式为 mm-dd-yyyy（月-日-年）。

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

比赛编号：一个 Wordle 谜题的索引，从 202 开始，日期为 2022 年 1 月 7 日。

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

词：解决方案单词，玩家们正在尝试猜测与相关日期和比赛编号相关的单词。

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

报告结果数量：当天在 Twitter 上记录的总得分。

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

困难模式下的分数：当天在 Twitter 上记录的困难模式分数数量。

1 try: The percentage of players solving the puzzle in one guess.

1 尝试：一次猜测解决难题的玩家百分比。

2 tries: The percentage of players solving the puzzle in two guesses.

2 次尝试：在两次猜测中解决难题的玩家百分比。

3 tries: The percentage of players solving the puzzle in three guesses.

3 次尝试：在三次猜测中解决难题的玩家百分比。

4 tries: The percentage of players solving the puzzle in four guesses.

4 次尝试：在四次猜测中解决难题的玩家百分比。

5 tries: The percentage of players solving the puzzle in five guesses.

5 次尝试：在五次猜测中解决难题的玩家百分比。

6 tries: The percentage of players solving the puzzle in six guesses.

6 次尝试：在六次猜测中解决难题的玩家百分比。

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.

7 次或更多尝试 (**X**)：未能在六次或更少尝试中解决难题的玩家百分比。注意：由于四舍五入，百分比可能并不总是加总为 100%。

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

2. Wordle 的玩法发布在纽约时报网站上。 ^[2]

How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word

Examples

W E A R Y

W is in the word and in the correct spot.

P I L L S

I is in the word but in the wrong spot.

V A G U E

U is not in the word in any spot



Glossary 术语表

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

纽约时报：一本总部位于美国纽约市的日报，提供印刷版和在线版。

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

Twitter：一个社交网络网站，允许用户发布不超过 280 个字符的短帖子（从最初的 140 个字符增加）。

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

解决（Wordle 拼图）：按正确顺序输入正确的字母，以形成当天的 Wordle 单词。

References 参考文献

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.

注意：我们提供以下引用以支持问题陈述。我们从这些资源中提取了重要的观点。解决这个 MCM 问题不需要这些网站上的额外信息。解决此问题不需要访问《纽约时报》或 Twitter 网站。

[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>.

[1] Wordle 标志来自《纽约时报》网站。访问于 2022 年 12 月 13 日，网址为 <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>.

[2] “Wordle-The New York Times.” 纽约时报, 2022. 访问于 2022 年 12 月 13 日，网址：<https://www.nytimes.com/games/wordle/index.html>。

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

[3] “Wordle-纽约时报。” 纽约时报，2022 年 7 月 21 日。

[4] “Wordle Stats.” Twitter, July 20, 2022.

[4] “Wordle Stats.” Twitter, 2022 年 7 月 20 日。