Wordle是一款近期很受欢迎的拼图游戏。该游戏的玩法是在限定次数对一个长度为五的单词猜测。每一次，玩家输入连续的五个字母，并被反馈告知输入的单词哪些存在并在正确的位置，哪些存在但在错误的位置，哪些不存在。MCM整理了该游戏有关每日单词，玩家数量，尝试次数的数据，

我们队伍将建立模型对这些数据分析，并完成以下工作：

建立模型拟合玩家人数随日期的变化，并对未来玩家人数进行预测。探究单词的各种属性对尝试次数百分比分布的影响。

预测一个给定的单词对尝试次数百分比分布的影响。

对现有单词按照难度分类并总结出每个类别单词的共同属性。评价给定单词的难度。

对报告结果的其他特征进行探索。

# Background

Wordle is a recently popular puzzle, which is played by guessing a word of length of five within a limited number of times. Each time, players type five letters in a row and are told by feedback which of the typed words were present and in the correct position, which were present but in the wrong position, and which were not present. MCM has compiled data about daily words, number of players, and percentage of each number of attempts for the game.



https://www.bbc.com/news/technology-60416057

# Problem Restatement

Based on the background information, our team will build a model to analyze these data and complete the following work:

* Build a model to fit the change of the number of players with the date, and predict the number of players in the future. Explore the influence of various attributes of words on the percentage distribution of attempts.
* Predict the effect of a given word on the percentage distribution of attempts.
* Classify the existing words according to the difficulty and summarize the common attributes of the words in each category. Rate the difficulty of a given word.
* Explore other features of the report results.

1. 假设玩家水平一样，这样总得分和总人数才能匹配。
2. 假设未来短期内游戏机制和宣传力度相同，这样曲线走向不会发生异常改变。
3. 假设游戏不存在作弊情况如刷分，交换答案等
4. 假设每天给定单词难度是随机的
5. 假设本文模型建立在我们规定的单词属性上