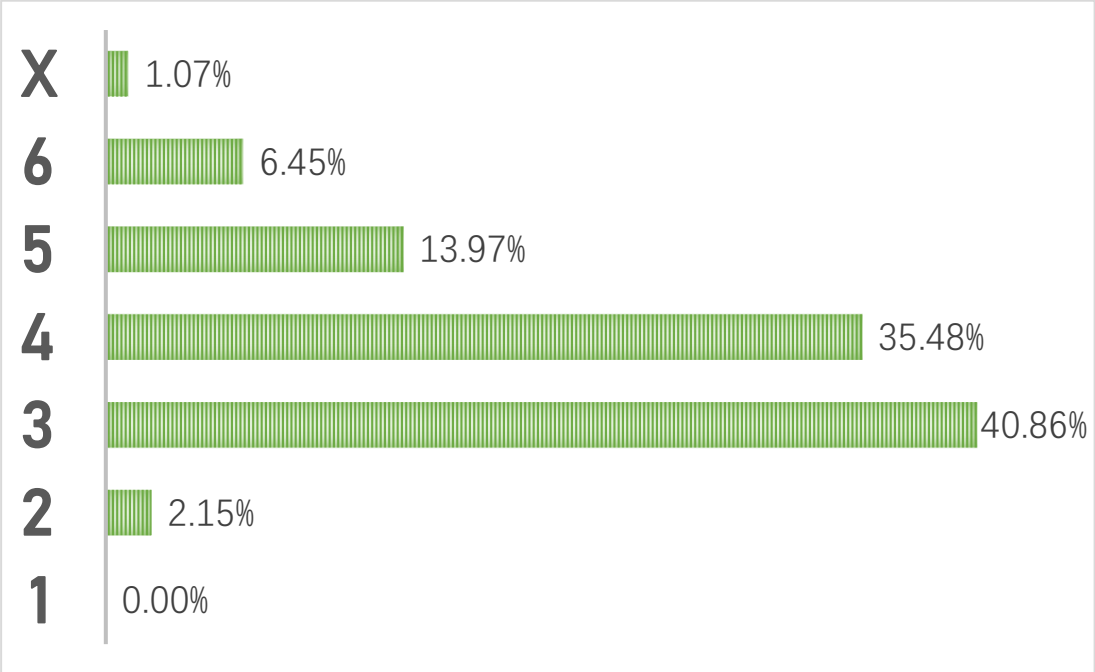
Dear Editor of New York Times:

We were asked to conduct an analysis of the results in the file and answer the questions you previously sent us, so we write to reply our result to you.

For the first question in which you asked us to create a prediction interval for reported result number as well as discuss the attributes of the word that may influence the percentage of players reported to play in Hard Mode, we notice that the curve of the numbers of reported results and hard mode are highly identical and so the correlation test has supported us. We use MLP(Multilayer Perceptron), which is a machine learning model, for prediction. The predicted result of the interval for reported result on March 1, 2023 is **[18450, 20490].** As for the attributes of the word, we take a Pearson Correlation Test on the number of repeated letters, the number of vowels, and the frequency of word appearing on Google and Twitter, finding that the percentage of players reported to play in Hard Mode is related to the frequency of the word with a P-value of 0.006, which is less than 0.05 and thus highly acceptable.

As for the second question, we were required to predict the distribution of reported results. So we use MLP and GBDT to construct model on the seven tries, using the output of the previous model as input of the next model to turn our model into memory model. It is worth mentioning that when we choose the input we use the result of either MLP or GBDT depending on their RMSEs. As a result, our prediction is based on both model. The prediction of word 'EERIE' is displayed in the picture below, with RMSEs of 0.2053 in minimum and 0.5937 on average.



For the third question, we should classify solutions by difficulty. We use K-means to divide all data points into 5 clusters and obtained the classification of each data point. Then we randomly split the dataset into training set and testing set in a 4:1

ratio. We used the data points in the training set as the original data, performed KNN classification on the testing set, obtained the test results and compared them with the testing set. The test accuracy rate is 100%. We classify words to five categories: Very Hard, Hard, Normal, Easy, Very Easy. According to K-Means, the word ‘eerie’ is categorized into ‘**Easy**’ Class.

We are also excited to share you with some interesting facts we found in the data. There are some words that are misspelling, such as ‘clean’ is written to ‘clen’ and ‘probe’ is written to ‘rprobe’. Besides, there exists a word ‘naïve’ in which ‘ï’ is unrecognizable by some algorithms. So we delete the wrong words and ‘naïve’.

Another interesting point lies in the fact that the percentage of Tries 7 decreases as time goes. If we compute an average value of Tries 7 every 7 rows, we can find that as time flies, people are more unlikely to end the game in 7 or more times.

We hope that our work will prove in every way satisfactory to you and if there is any other particulars required, we shall be pleased to give you.

Yours sincerely