The New York Times introduced a web-based word game called Wordle in 2022. The game is to guess a five-letter word in six tries or fewer. However, the data from January 1, 2022 to December 31, 2022 shows that the numbers of reported players continued to decline after a short-term increase. Therefore, through the analysis of historical data, this report discovered some potential patterns and characteristics for the game to enjoy long term prosperity.

We first performed data cleaning and adjustments followed by analyzing the relation of attributes and times. We conduct time series analysis and natural language processing methods to retrieve 82 time-series features and 47 word-related features. The goal of our model is to find out number of reports’s envolving trend with time, the relationship between the word attribute and the distribution of tries, and predict the distribution guessed at a given day. Specifically, the study includes the following parts.

Firstly, to predict the number of reported results, we used GluonTS and ARIMA to make prediction of March 1st and compared the models; based on those, we successfully provided a prediction interval. We also find a strong correlation between hard mode percentage and the time while the goal is to predict its relationship with word features. Therefore, we picked a less correlated interval, and use different machine-learning algorithms as models.

Secondly, to predict the distribution of the reported results on March 1st with the word “EERIE”, we use both time series and word-related features and apply machine learning models containing XGBoost, Random Forest, and Tree Algorithms to train the data. The label of difficulty levels was derived using the K-means algorithm to classify the data into three clusters of easy, medium and hard mode. We also apply the machine learning model to predict the classification and we achieved an accuracy rate of more than 70% and successfully classified the word "EERIE" as a hard difficulty level, which confirmed our assumption.

Based on our predictions and data analysis, we wrote a letter of suggestions to the New York Times, analyzing the potential dangers of the game and offering a few suggestions for future game development. Our suggestions included introducing new features to the game, making the game more accessible to non-English speakers, and introducing a more comprehensive scoring system to encourage players to play the game in different ways. In summary, our model uses various data processing techniques to analyze the Wordle game’s data and discover its features. Our predictions and suggestions can help the New York Times improve the game and attract more players to the game.