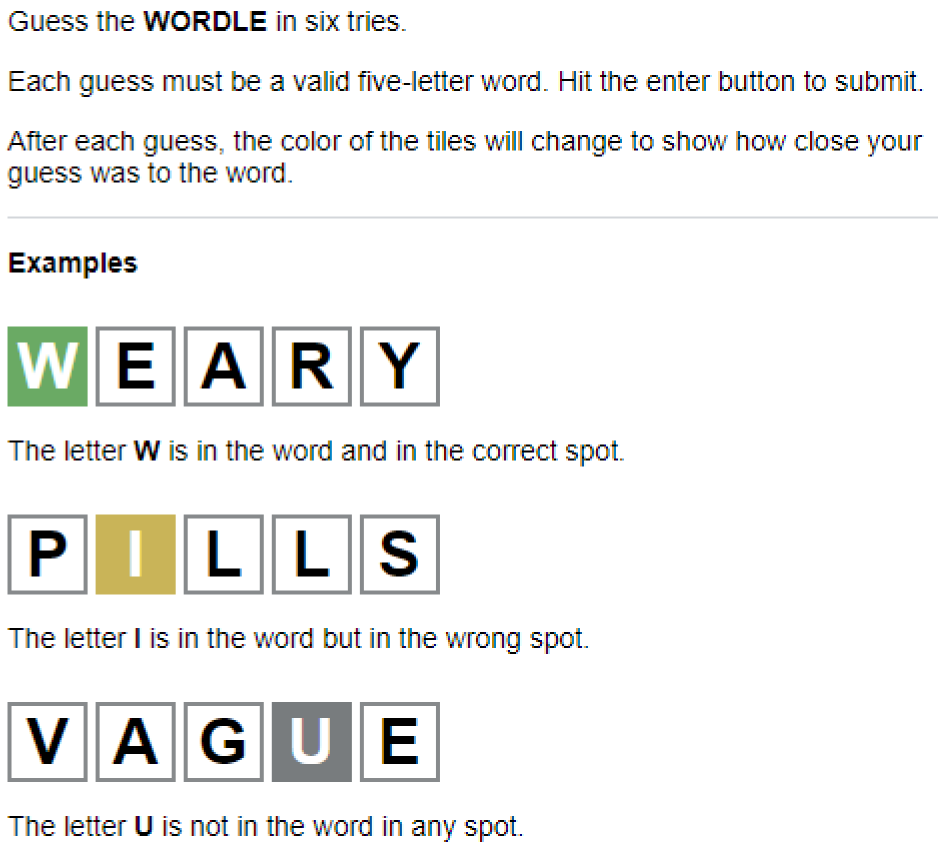
What is Wordle?

WORDLE

Wordle is a daily online word game created by Josh Wardle and later sold to The New York Times Company. Every 24 hours, there is a new word of the day, and we have to figure it out in 6 tries. Since its launch, the game has become very popular. According to The New Times, over 300,000 people play the game daily.

How to play?



Guess the **WORDLE** in six tries.

Each guess must be a valid five-letter word. Hit the enter button to submit.

After each guess, the color of the tiles will change to show how close your guess was to the word.

Examples

W E A R Y

The letter **W** is in the word and in the correct spot.

P I L L S

The letter **I** is in the word but in the wrong spot.

V A G U E

The letter **U** is not in the word in any spot.

Image source:https://www.nytimes.com/games/wordle/index.htm

Problem statement:

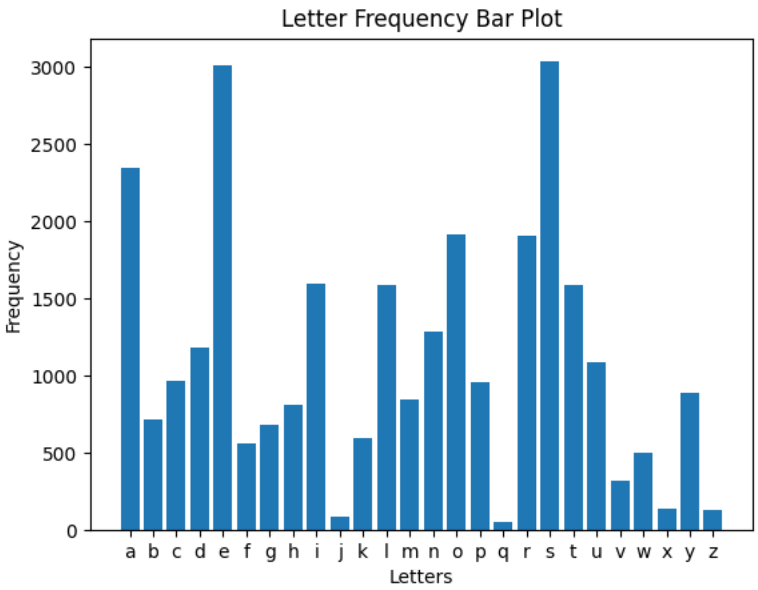
Find a strategy to guess the right word in Wordle in the least number of guesses.

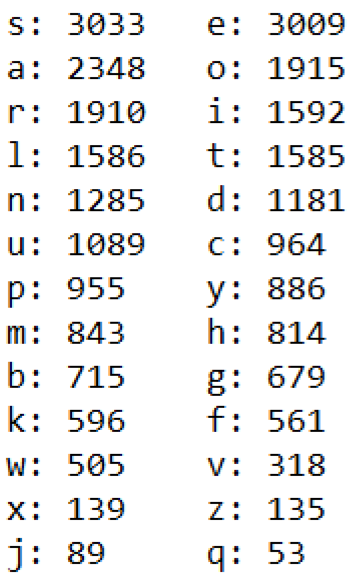
Methodology:

1. Downloading a database of 5-letter words.
2. Analyse these words
3. Propose a strategy based on the analysis
4. Testing our strategy.
    a. Write a program to guess the correct word using our strategy.
    b. Apply the program for every word in the database and take the average number of guesses.
    c. Develop a program to apply an algorithm on the Wordle website to guess the word.

Analysis

Used Python to analyse a database of 5760 5-letter words. Then, this data was used to plot graphs, which helped further understand the data.





```
s: 3033   e: 3009
a: 2348   o: 1915
r: 1910   i: 1592
l: 1586   t: 1585
n: 1285   d: 1181
u: 1089   c: 964
p: 955    y: 886
m: 843    h: 814
b: 715    g: 679
k: 596    f: 561
w: 505    v: 318
x: 139    z: 135
j: 89     q: 53
```

Here, the frequency of each letter occurring in all words.

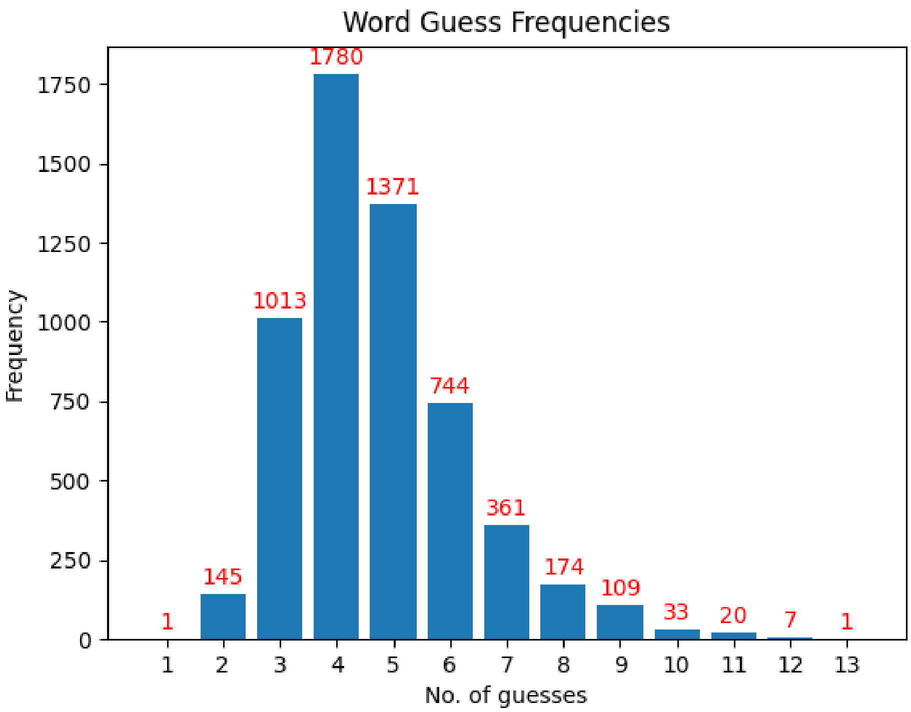
Strategy

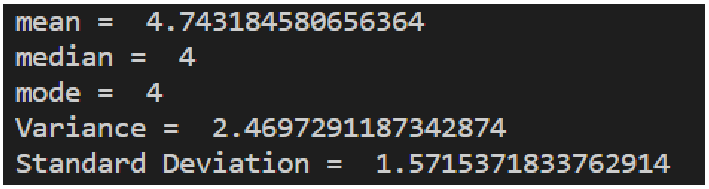
Based on the analysis done on our database, I developed strategies to guess the right word in the least number of guesses and within 6 guesses. First, pass fixed words as initial guesses and then the code will find the correct position, wrong positions and letters not present in the word. Based on those details, the code will decide the next guess.

Method -1:

Let's choose the first 5 most occurring letters: [S, E, A, O, R]
Make one word from it: {AROSE}
So, the first initial guess will be the word 'AROSE' each time, and then the code will provide further guesses.

Result: After applying this algorithm for all words, the output is as follows.





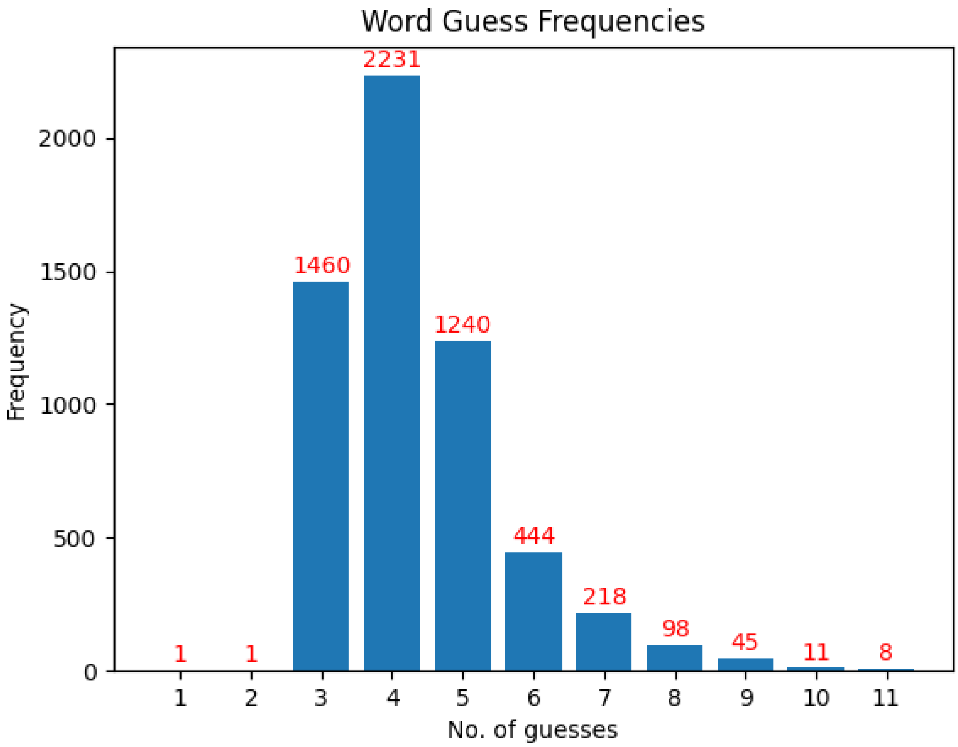
```
mean =  4.743184580656364
median =  4
mode =  4
Variance =  2.4697291187342874
Standard Deviation =  1.5715371833762914
```

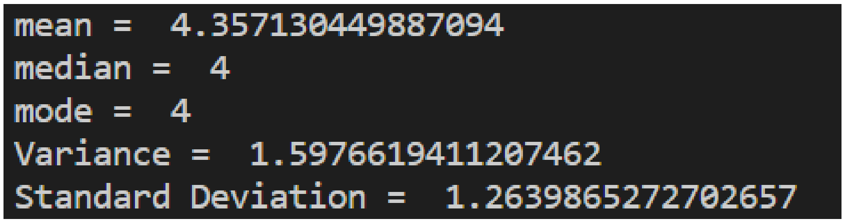
But we can see here that 705 words are guessed at an attempt larger than 6.

Method -2:

Let's choose the first 10 most occurring letters: [S, E, A, O, R, I, L, T, N, D]
Make two words from it: {STAND, OILER}
So, the first initial guess will be the word ‘STAND’, and the second guess will be the word ‘OILER’ each time, and then the code will provide further guesses.

Result: After applying this algorithm for all words, the output is as follows.





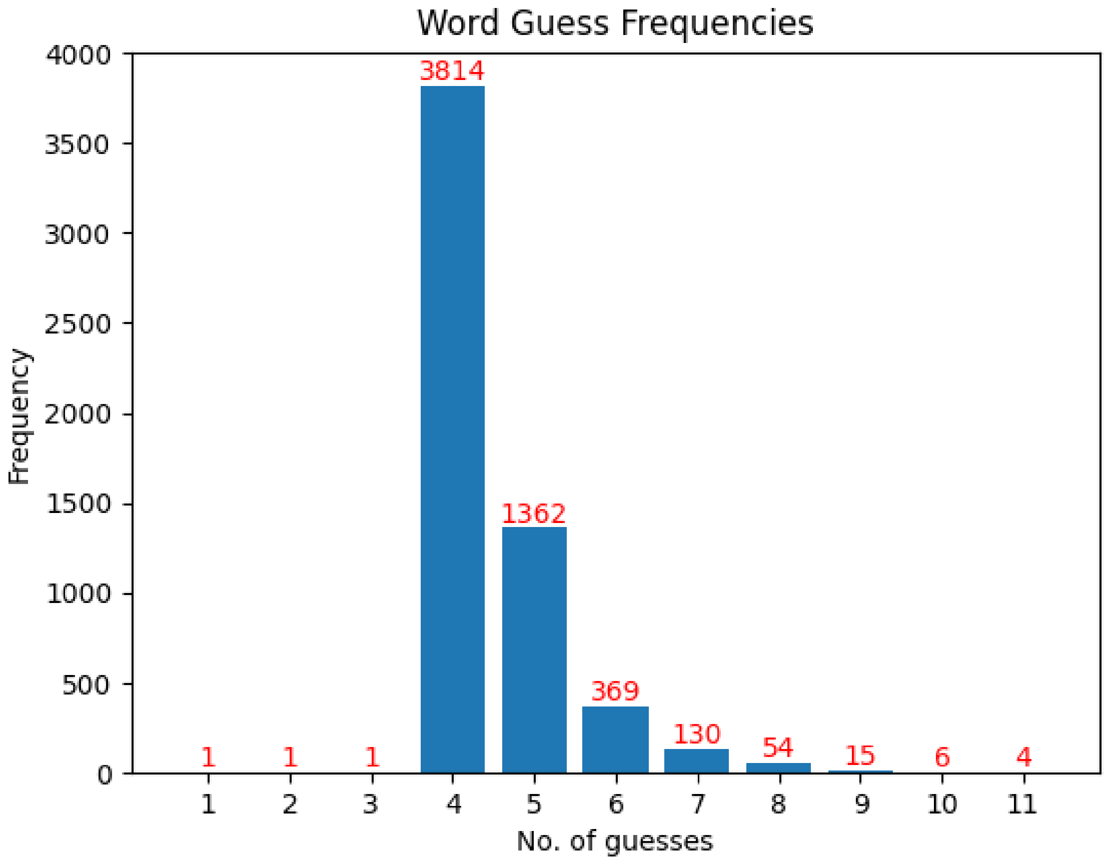
```
mean =  4.357130449887094
median =  4
mode =  4
Variance =  1.5976619411207462
Standard Deviation =  1.2639865272702657
```

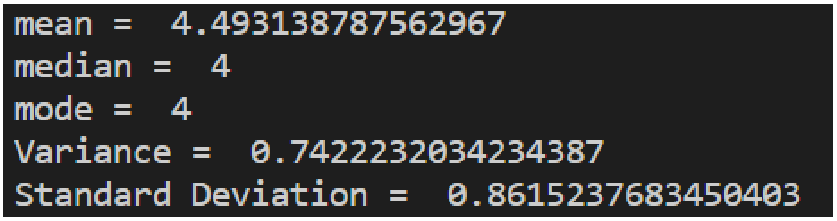
But we can see that 380 words are guessed at an attempt larger than 6.

Method -3:

Let's choose the first 15 most occurring letters: [S, E, A, O, R, I, L, T, N, D, U, C, P, Y, M]
Make three words from it: {CANDY, PLUME, RIOTS}
So, the first initial guess will be the word ‘CANDY’, the second guess will be the word ‘PLUME’, and the third guess will be the word ‘RIOTS’ each time, and then the code will provide further guesses.

Result: After applying this algorithm for all words, the output is as follows.





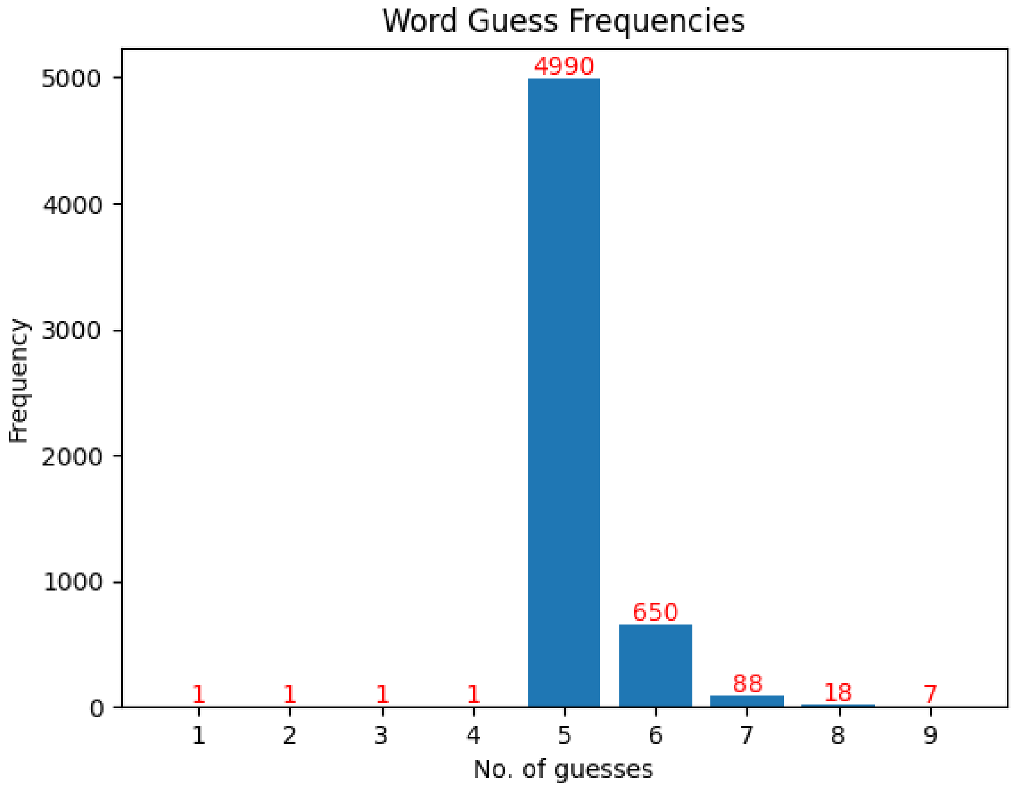
```
mean =  4.493138787562967
median =  4
mode =  4
Variance =  0.7422232034234387
Standard Deviation =  0.8615237683450403
```

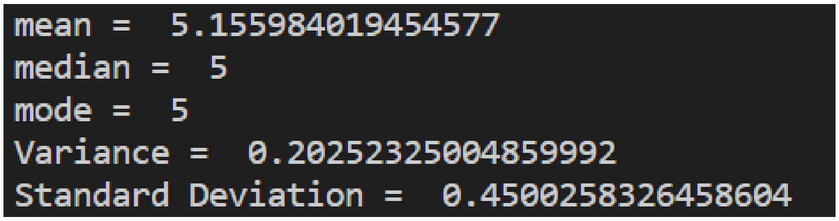
But we can see that 209 words are guessed at an attempt larger than 6.

Method -4:

Let's choose the first 20 most occurring letters:
[S, E, A, O, R, I, L, T, N, D, U, C, P, Y, M, H, B, G, K, F]
Make four words from it: {SHELF, DIRTY, BACON, GRUMP}
So, the first initial guess will be the word ‘SHELF’, the second guess will be the word ‘DIRTY’, the third guess will be the word ‘BACON’, and the fourth guess will be the word ‘GRUMP’ each time, and then the code will provide further guesses.

Result: After applying this algorithm for all words, the output is as follows.





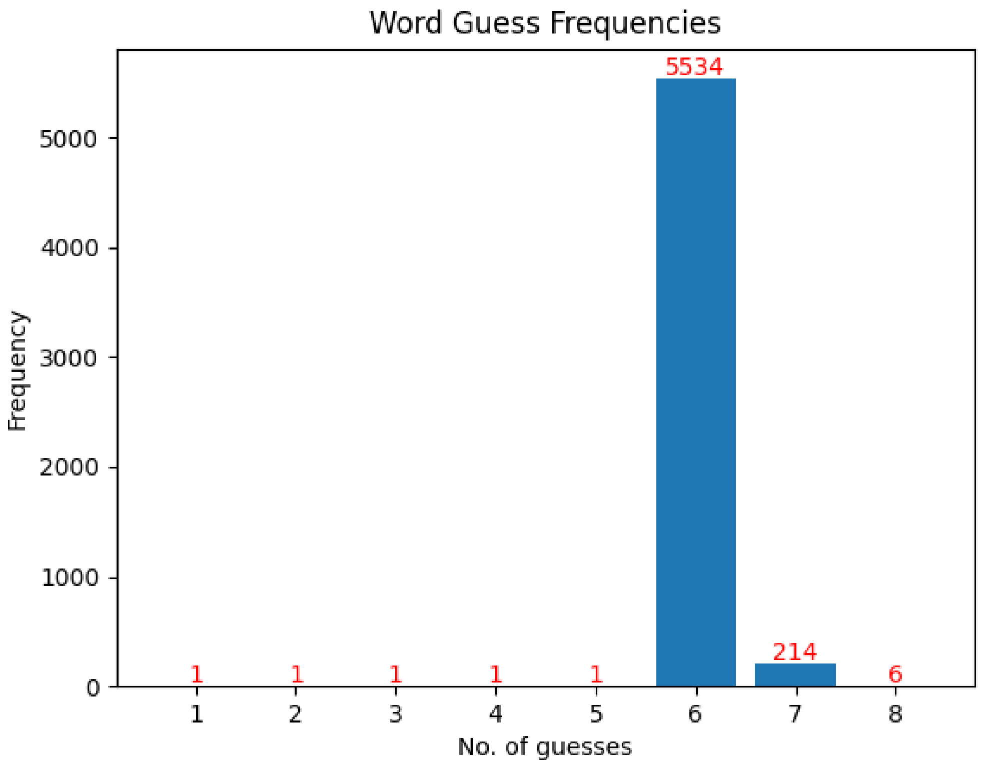
```
mean =  5.155984019454577
median =  5
mode =  5
Variance =  0.20252325004859992
Standard Deviation =  0.4500258326458604
```

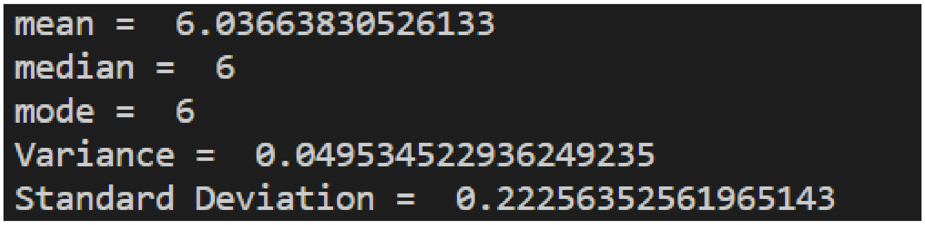
But we can see that 113 words are guessed at an attempt larger than 6.

Method -5:

Let's choose the first 25 most occurring letters:
[S, E, A, O, R, I, L, T, N, D, U, C, P, Y, M, H, B, G, K, F, W, V, X, Z, J]
Make four words from it: {FJORD, GUCKS, NYMPH, VIBEX, WALTZ}
So, the first initial guess will be the word ‘FJORD’, the second guess will be the word ‘GUCKS’, the third guess will be the word ‘NYMPH’, the fourth guess will be the word ‘VIBEX’, and the fifth guess will be the word ‘WALTZ’ each time, and then the code will provide further guesses.

Result: After applying this algorithm for all words, the output is as follows.



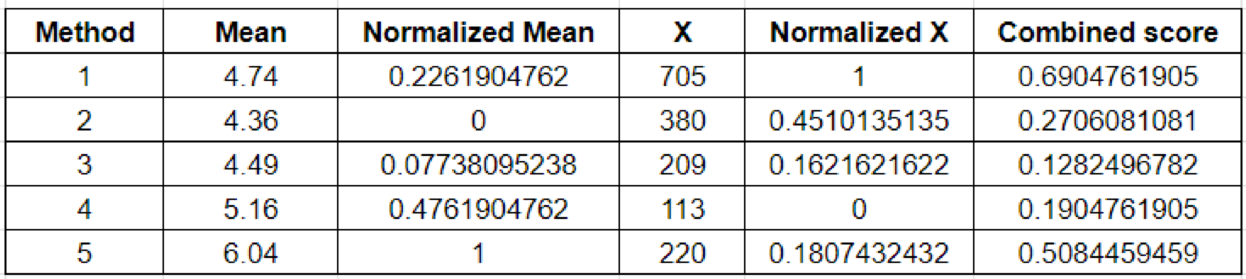


```
mean =  6.03663830526133
median =  6
mode =  6
Variance =  0.049534522936249235
Standard Deviation =  0.22256352561965143
```

But we can see that 220 words are guessed at an attempt larger than 6.

Conclusion:

X = no. of guesses greater than 6



Method	Mean	Normalized Mean	X	Normalized X	Combined score
1	4.74	0.2261904762	705	1	0.6904761905
2	4.36	0	380	0.4510135135	0.2706081081
3	4.49	0.07738095238	209	0.1621621622	0.1282496782
4	5.16	0.4761904762	113	0	0.1904761905
5	6.04	1	220	0.1807432432	0.5084459459

Here, first normalized both the data columns using max-min normalisation. Then, for the final combined score, selected weightage of 60% for the value of X and 40% for the value of a mean as we want to minimise guess number and mean also, but priority is given to X value.

$$\boxed{Normalized\ value\ =\ \frac{X_i - \mu}{\sigma}}$$

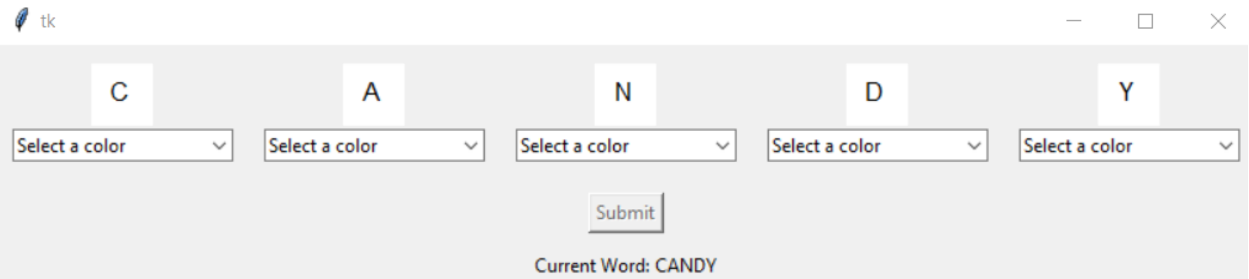
where μ = mean and σ = standard deviation

$$\boxed{Combined\ Score\ =\ (0.4\ *\ Normalized\ mean)\ +\ (0.6\ *\ Normalized\ X)}$$

So according to calculation Method-3 has less value for a combined score, hence it is the most optimised solution.

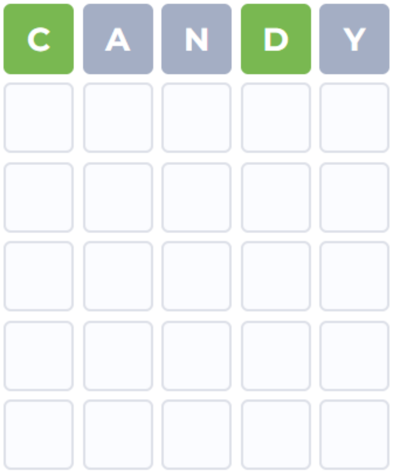
Python Program working

Step - 1: The first code will give an initial guess of ‘ CANDY.’

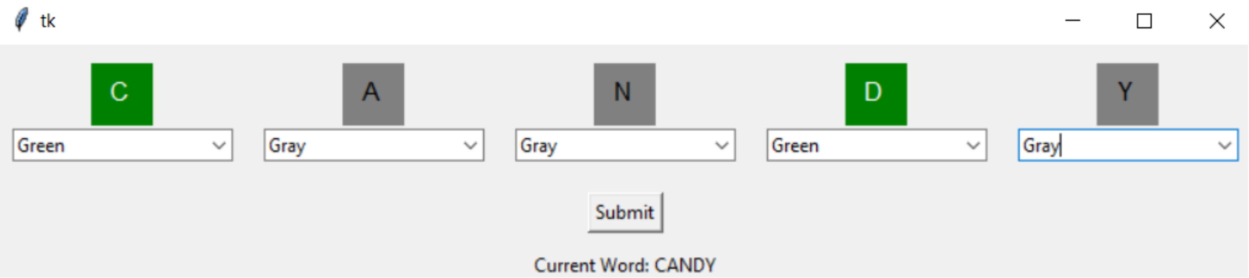


Current Word: CANDY

Step - 2: Then we will enter the word ‘ CANDY ’ on the webpage of the WORDLE game.



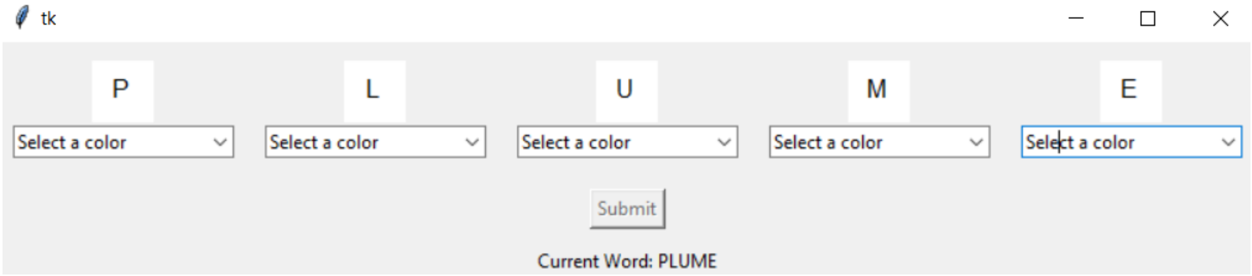
Step - 3: The respective colours we will fill in the Python GUI window.



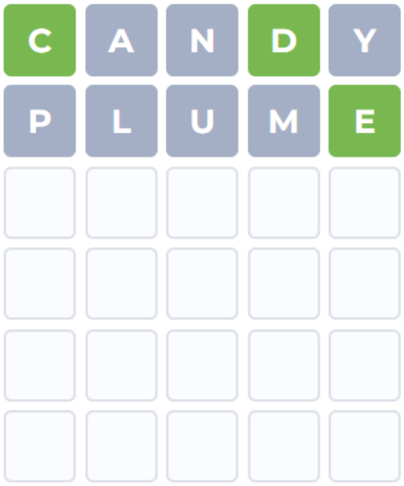
C: Green
A: Gray
N: Gray
D: Green
Y: Gray

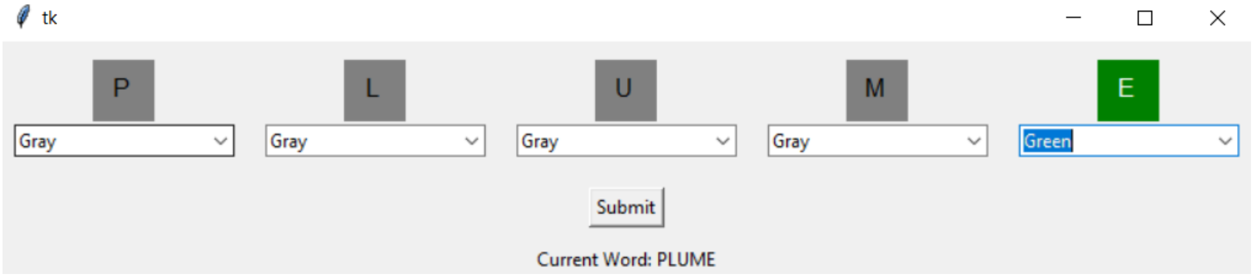
Current Word: CANDY

Step - 4: After clicking on the ‘Submit’ button, the code will give us the next guess and this process continues until a word is guessed.



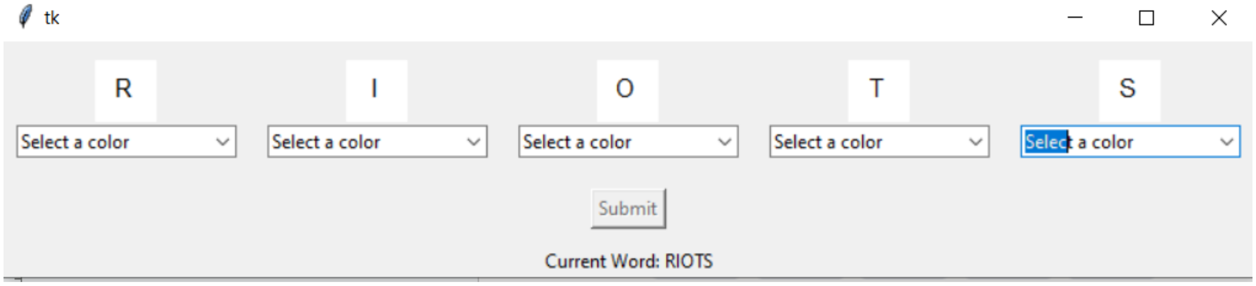
Current Word: PLUME



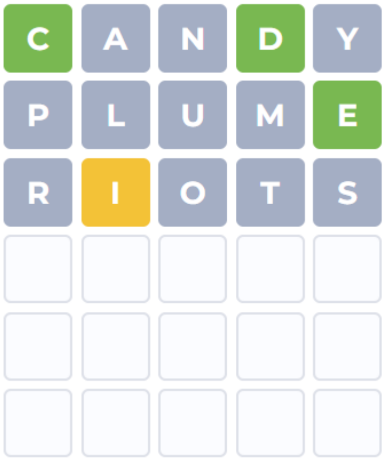


Gray Gray Gray Gray Green

Current Word: PLUME



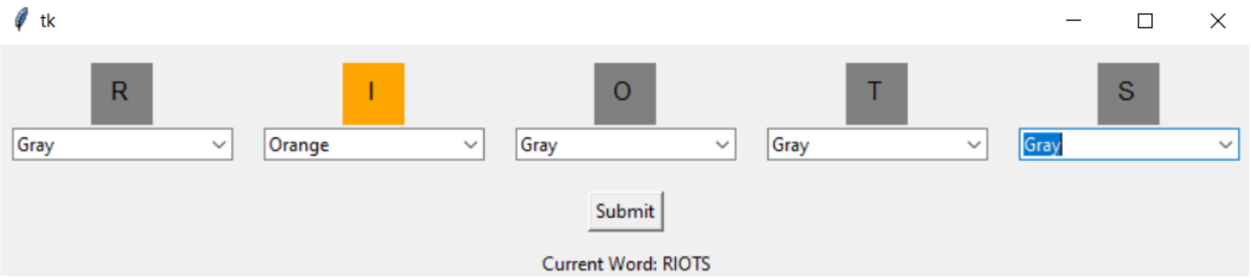
Current Word: RIOTS



C A N D Y

P L U M E

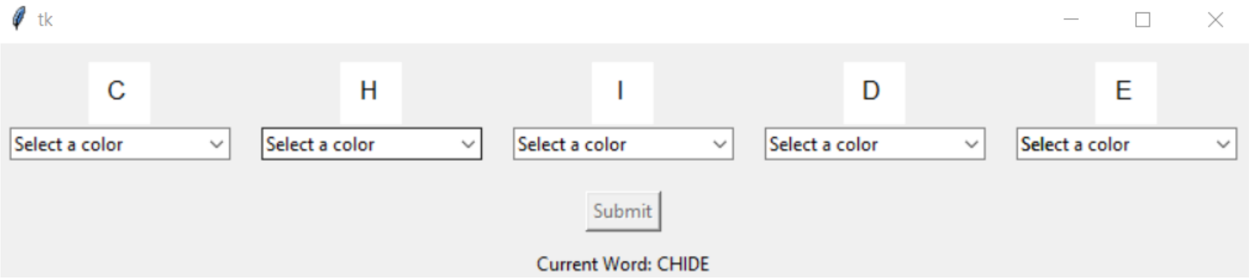
R I O T S



R I O T S

Gray | Orange | Gray | Gray | Gray

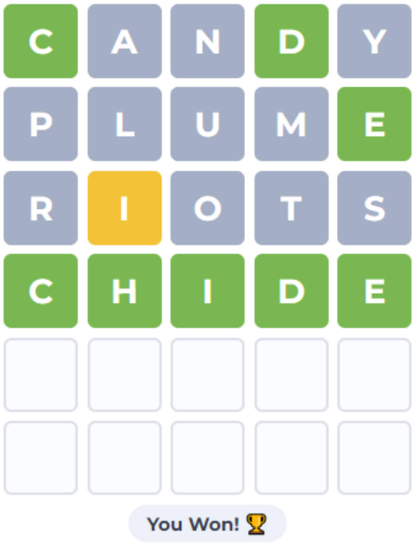
Current Word: RIOTS



C H I D E

Select a color | Select a color | Select a color | Select a color | Select a color

Current Word: CHIDE



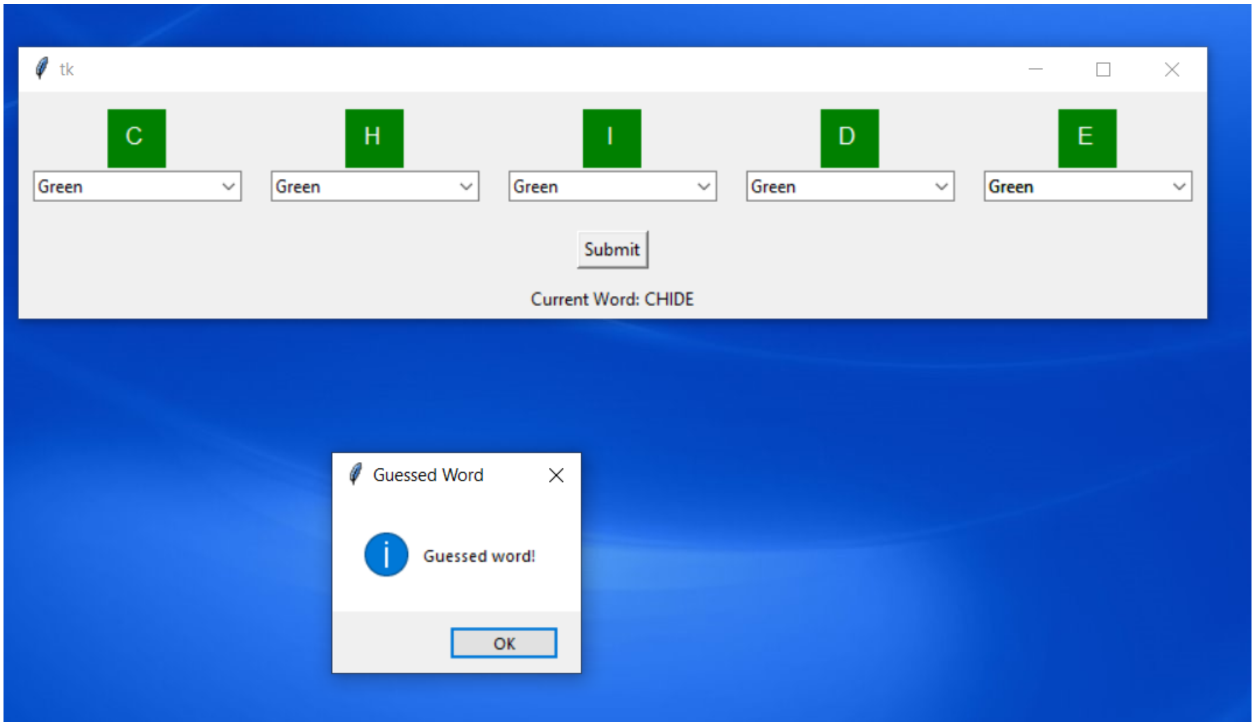
C A N D Y

P L U M E

R I O T S

C H I D E

You Won! 🏆



```
Green          Green          Green          Green          Green

Current Word: CHIDE
```

```
Guessed Word

Guessed word!
```