

WORDLE

Group Members and Roles:

Ananna Biswas - *Test Engineer*

Karthik Garimella - *Programmer*

Riya Mole - *Documentation and Reviewer*

Sandeep Alfred - *Algorithm Designer*

Objective: To build a naive agent to guess a word from a dataset in six attempts.

Game rules:

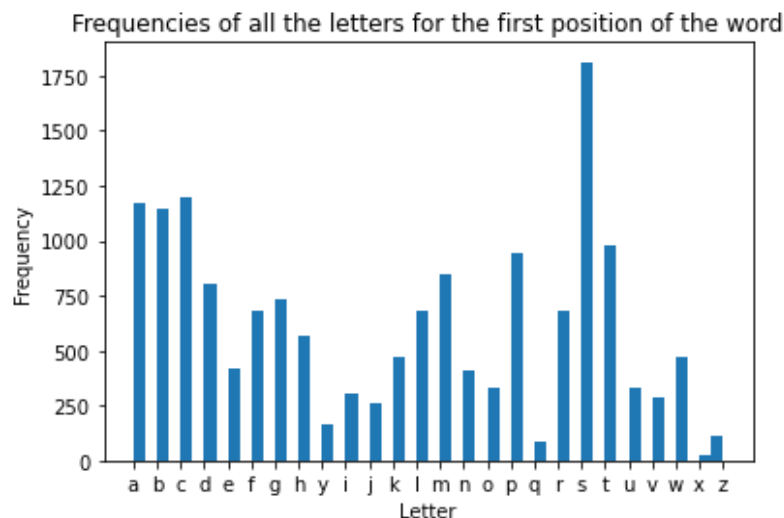
1. The objective is to guess a target word, which is a five-letter word.
2. Six attempts to guess the target word correctly.
3. After each guess, feedback is received in the form of colored tiles for each letter for the guessed word:
 - Green: The letter is correct and in the correct position.
 - Yellow: The letter is correct but in the wrong position.
 - Red*: The letter is incorrect and not in the word.

*Since we could not visualize the color Gray in the terminal we decided to choose the color Red to denote the letter is incorrect and does not exist in the word.

Our Approach:

1.) Analysis:

- We analyzed the words dataset words_alpha.txt. Our findings saw specific letters to occur more frequently in certain positions.
 - Example: The following graph shows the frequencies of all the letters for the first position of the word.



2.) Designing the dataframe:

- We decided to calculate the total frequency of each word and sort the dataframe with the word which has the highest cumulative frequency.
 - The dataframe with the calculated weights is shown below.

	words	letters	letter_1	letter_2	letter_3	letter_4	letter_5	letter_1_freq	letter_3_freq	letter_4_freq	letter_2_freq	letter_5_freq	CumiWeight
0	sanes	[s, a, n, e, s]	s	a	n	e	s	1813	1238	2510	2871	3148	11580
1	sales	[s, a, l, e, s]	s	a	l	e	s	1813	1061	2510	2871	3148	11403
2	sores	[s, o, r, e, s]	s	o	r	e	s	1813	1545	2510	2281	3148	11297
3	cares	[c, a, r, e, s]	c	a	r	e	s	1196	1545	2510	2871	3148	11270
4	bares	[b, a, r, e, s]	b	a	r	e	s	1141	1545	2510	2871	3148	11215
...
15915	ewhow	[e, w, h, o, w]	e	w	h	o	w	421	208	903	174	94	1800
15916	iddhi	[i, d, d, h, i]	i	d	d	h	i	301	514	288	136	509	1748
15917	zmudz	[z, m, u, d, z]	z	m	u	d	z	112	787	545	233	54	1731
15918	enzym	[e, n, z, y, m]	e	n	z	y	m	421	143	161	557	297	1579
15919	oghuz	[o, g, h, u, z]	o	g	h	u	z	334	208	686	102	54	1384

15920 rows × 13 columns

- The `initialise_dataframe(wordle_list)` function creates the dataframe and assigns it to the `DF_GUESS` global variable for easy access across the program.
- The `calculate_weights()` function calculates the weights of the newly generated dataframe each time changes are made to it.

3.) Decision-making logic:

- Based on the feedback received after each guess the `guess_word(feedback, previousGuess, gs_ys)` function updates the `DF_GUESS` dataframe with the new possible words.
- `feedback` argument is a list of letters with either G, R, or Y for each position of the `previousGuess`. Example: ['G', 'G', 'R', 'Y', 'R']
- We iterate through the feedback received for each letter of the previous guess, `focus_letter` denotes that letter.
- If the `focus_letter` color is green, the dataframe is updated with all the words having the `focus_letter` in the current position of the iteration.
- If the `focus_letter` color is yellow, the dataframe is updated with all the words containing the `focus_letter`, and all the words containing the focus letter in the current position are removed from the dataframe.
- If the `focus_letter` color is red and the letter is also marked green or yellow, all the words containing the `focus_letter` only in the current position are deleted. Otherwise, if the `focus_letter` is red and is not marked green or yellow, all the words containing the `focus_letter` are deleted.
- Weights are again calculated for the newly updated dataframe which is based on the previous guess.

3.) Testing the agent:

- The code logic was tested for all the values of the dataset. The agent was able to guess the words in 6 attempts 80% of the time. The range of guesses lies in the interval 1 to 15.
 - Following is the analysis of the test data.

