WordleStuff

February 4, 2022

```
In [235]: import pandas as pd
```

1 Assessing Potential Answers

All spreadsheets were created by Zach Wissner-Gross, who provided them via Google Sheets. (Links found here: https://fivethirtyeight.com/features/when-the-riddler-met-wordle/and via Jeremy Vine's Data is Plural mailing list.)

```
In [236]: answers = pd.read_csv('MysteryWords.csv', squeeze=True)

In [237]: answers.head()

Out[237]: 0    abase
          1    abate
          2    abbey
          3    abbot
          4    abhor
          Name: aback, dtype: object
```

Of all potential answers, first I want to see what the most common vowels are, and the most common consonants.

```
In [238]: allAnswers = ''.join(answers.tolist())

In [239]: all_freq = {}

          for i in allAnswers:
              if i in all_freq:
                  all_freq[i] += 1
              else:
                  all_freq[i] = 1

In [240]: all_freq

Out[240]: {'a': 977,
           'b': 280,
           's': 669,
           'e': 1233,
```

```
 't': 729,
 'y': 425,
 'o': 754,
 'h': 389,
 'r': 899,
 'i': 671,
 'd': 393,
 'l': 719,
 'u': 467,
 'v': 153,
 'c': 476,
 'n': 575,
 'g': 311,
 'p': 367,
 'm': 316,
 'f': 230,
 'x': 37,
 'w': 195,
 'k': 209,
 'z': 40,
 'j': 27,
 'q': 29}
```

```
In [241]: allFreqDF = pd.DataFrame([all_freq])

In [242]: allFreqDF
```

Out[242]:

	a	b	c	d	e	f	g	h	i	j	...	q	r	s	t	\
0	977	280	476	393	1233	230	311	389	671	27	...	29	899	669	729	

	u	v	w	x	y	z
0	467	153	195	37	425	40

[1 rows x 26 columns]

```
In [243]: vowels = allFreqDF[["a","e","i","o","u"]]

In [244]: vowels.sort_values(by=0, ascending=False, axis=1)
```

Out[244]:

	e	a	o	i	u
0	1233	977	754	671	467

```
In [245]: consonants = allFreqDF.drop(["a","e","i","o","u"],axis=1)

In [246]: consonants.sort_values(by=0, ascending=False, axis=1)
```

Out[246]:

	r	t	l	s	n	c	y	d	h	p	...	g	b	f	k	\
0	899	729	719	669	575	476	425	393	389	367	...	311	280	230	209	

```
     w    v   z   x   q   j
0  195  153  40  37  29  27

[1 rows x 21 columns]
```

So now we know that the two most common vowels are e and a, and the four most common consonants are r,t,l and s. (Numbers chosen somewhat at random, but on the theory that every word is going to have at least 1 vowel/4 consonants, and maybe 2 vowels/3 consonants.

```
In [247]: answersDF = []
          answersDF = pd.DataFrame(answersDF, columns=['words'])
          answersDF["words"] = answers

In [248]: answersDF["commonality_score"] = 0
```

Next we'll assign a commonality score, to find out which words have the most common letters.

```
In [249]: for index,row in answersDF.iterrows():
              current_word = str(row['words'])

              if 'a' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s

              if 'e' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s

              if 'r' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s

              if 't' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s

              if 'l' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s

              if 's' in current_word:
                  answersDF.at[index, 'commonality_score'] = answersDF.at[index, 'commonality_s


In [250]: answersDF.sort_values(by=['commonality_score'],ascending=False).head(10)

Out[250]:       words  commonality_score
          66    alter                  5
          1776  slate                  5
          1901  stale                  5
          47    alert                  5
          1101  later                  5
          1907  stare                  5
          1915  steal                  5
```

```
1114  least                5
1914  steak                4
1909  start                4
```

And so we get a working list of the words that are most likely to have common letters!