

```
In [1]: import pandas as pd
```

```
In [2]: df = pd.read_csv('words.csv', index_col='Word')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	Char Count	Value
Word		
aa	2	2
aah	3	10
aahed	5	19
aahing	6	40
aahs	4	29

## ▼ Activities

### ▼ How many elements does this dataframe have?

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 172821 entries, aa to zyzyvas
Data columns (total 2 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Char Count  172821 non-null  int64
1   Value       172821 non-null  int64
dtypes: int64(2)
memory usage: 4.0+ MB
```

```
In [6]: df.shape
```

```
Out[6]: (172821, 2)
```

### ▼ What is the value of the word *microspectrophotometries* ?

```
In [9]: df['Value'].loc['microspectrophotometries']
```

```
Out[9]: 317
```

```
In [10]: df.loc['microspectrophotometries', 'Value']
```

```
Out[10]: 317
```

df.loc[Index, Columns]

▼ **What is the highest possible value of a word?**

In [11]: df['Value'].max()

Out[11]: 319

In [12]: df.max()

Out[12]: Char Count 28  
Value 319  
dtype: int64

In [13]: df.describe()

Out[13]:

	Char Count	Value
count	172821.000000	172821.000000
mean	9.087628	107.754179
std	2.818285	39.317452
min	2.000000	2.000000
25%	7.000000	80.000000
50%	9.000000	103.000000
75%	11.000000	131.000000
max	28.000000	319.000000

▼ **Which of the following words have a Char Count of 7 and a Value of 87 ?**

In [14]: df.loc[['pinfish', 'glowing', 'enfold', 'microbrew']]

Out[14]:

	Char Count	Value
pinfish	7	81
glowing	7	87
enfold	6	56
microbrew	9	106

▼ **What is the highest possible length of a word?**

```
In [15]: df['Char Count'].max()
```

```
Out[15]: 28
```

▼ **What is the word with the value of 319 ?**

```
In [20]: filt = (df['Value'] == 319)
df[filt]
# or df.loc[filt]
```

```
Out[20]:
```

	Char Count	Value
<b>Word</b>		
reinstitutionalizations	23	319

```
In [24]: df.sort_values(by='Value', ascending=False)
```

```
Out[24]:
```

	Char Count	Value
<b>Word</b>		
reinstitutionalizations	23	319
microspectrophotometries	24	317
microspectrophotometry	22	309
microspectrophotometers	23	308
immunoelectrophoretically	25	307
...	...	...
aba	3	4
baa	3	4
ab	2	3
ba	2	3
aa	2	2

172821 rows × 2 columns

▼ **What is the most common value?**

In [26]: `df['Value'].value_counts()`

Out[26]:

Value	count
93	1965
100	1921
95	1915
99	1907
92	1902
...	
287	1
291	1
294	1
5	1
278	1

Name: count, Length: 303, dtype: int64

▼ **What is the shortest word with value 274 ?**

In [28]: `filt = (df['Value'] == 274)`  
`df[filt].sort_values(by='Char Count')`

Out[28]:

Word	Char Count	Value
overprotectivenesses	20	274
countercountermeasure	21	274
psychophysiologically	21	274

In [38]: `filt = (df['Value'] == 274)`  
`filt2 = (df['Char Count'] == df.loc[filt, 'Char Count'].min())`  
`df.loc[filt&filt2]`

Out[38]:

Word	Char Count	Value
overprotectivenesses	20	274

▼ **Create a column *Ratio* which represents the 'Value Ratio' of a word**

```
In [41]: df['Ratio'] = df['Value'] / df['Char Count']
df.head()
```

```
Out[41]:
```

	Char Count	Value	Ratio
Word			
aa	2	2	1.000000
aah	3	10	3.333333
aahed	5	19	3.800000
aahing	6	40	6.666667
aahs	4	29	7.250000

▼ **What is the maximum value of Ratio ?**

```
In [42]: df['Ratio'].max()
```

```
Out[42]: 22.5
```

▼ **What word is the one with the highest Ratio ?**

```
In [44]: filt = (df['Ratio'] == df['Ratio'].max())
df[filt]
```

```
Out[44]:
```

	Char Count	Value	Ratio
Word			
xu	2	45	22.5

```
In [47]: df.sort_values(by='Ratio', ascending=False).head(1)
```

```
Out[47]:
```

	Char Count	Value	Ratio
Word			
xu	2	45	22.5

▼ **How many words have a Ratio of 10 ?**

```
In [62]: filt = (df['Ratio'] == 10)
df[filt].shape
```

```
Out[62]: (2604, 3)
```

```
In [64]: df.query('Ratio == 10').shape
```

```
Out[64]: (2604, 3)
```

In [65]: `df['Ratio'].value_counts()`

Out[65]:

Ratio	
12.000000	3751
11.000000	3428
13.000000	3272
10.000000	2604
14.000000	2357
...	
10.550000	1
8.944444	1
8.941176	1
9.263158	1
21.250000	1

Name: count, Length: 1333, dtype: int64

▼ **What is the maximum Value of all the words with a Ratio of 10 ?**

In [69]: `filt = (df['Ratio'] == 10)`  
`df[filt].sort_values(by='Value', ascending=False)`

Out[69]:

	Char Count	Value	Ratio
Word			
electrocardiographically	24	240	10.0
electroencephalographies	24	240	10.0
electroencephalographer	23	230	10.0
phonocardiographic	18	180	10.0
inconceivabilities	18	180	10.0
...	...	...	...
web	3	30	10.0
bug	3	30	10.0
elm	3	30	10.0
as	2	20	10.0
oe	2	20	10.0

2604 rows × 3 columns

In [71]: `filt = (df['Ratio'] == 10)`  
`df.loc[filt, 'Value'].max()`

Out[71]: 240

▼ **Of those words with a Value of 260 , what is the lowest Char Count found?**

```
In [74]: filt = (df['Value'] == 260)
df.loc[filt, 'Char Count'].min()
```

Out[74]: 17

▼ **Based on the previous task, what word is it?**

```
In [78]: filt = (df['Value'] == 260)
filt2 = (df['Char Count'] == df.loc[filt, 'Char Count'].min())
df.loc[filt&filt2]
```

Out[78]:

	Char Count	Value	Ratio
hydroxytryptamine	17	260	15.294118

▼ **Find all the words with Char Count > avg Char Count**

```
In [84]: mean_char_count = df['Char Count'].mean()

filt = (df['Char Count'] > mean_char_count)
df[filt]
```

Out[84]:

	Char Count	Value	Ratio
aardwolves	10	120	12.000000
abacterial	10	72	7.200000
abandoners	10	93	9.300000
abandoning	10	81	8.100000
abandonment	11	103	9.363636
...	...	...	...
zygomorphies	12	176	14.666667
zygomorphy	10	168	16.800000
zygosities	10	154	15.400000
zygospores	10	165	16.500000
zymologies	10	146	14.600000

67582 rows × 3 columns

