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

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

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

Out[3]:

Word	Char Count	Value
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 zyzzyvas
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 [5]:
```
df.shape
```

Out[5]:
```
(172821, 2)
```

What is the value of the word `microspectrophotometries` ?

In [7]:
```
df.loc["microspectrophotometries","Value"]
```

Out[7]:
```
317
```

What is the highest possible value of a word?

In [8]:
```
df["Value"].max()
```

Out[8]:
```
319
```

In [9]:
```
df.max()
```

Out[9]:
```
Char Count     28
Value         319
dtype: int64
```

In [11]:
```
df.describe()
```

Out[11]:

	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 [13]:
```
df.loc[["superheterodyne","microbrew","glowing","enfold","pinfish"],"Value"]
```

Out[13]:
```
Word
superheterodyne    198
microbrew          106
glowing             87
enfold              56
pinfish             81
Name: Value, dtype: int64
```

In [14]:
```
df.loc[["superheterodyne","microbrew","glowing","enfold","pinfish"]]
```

Out[14]:

Word	Char Count	Value
superheterodyne	15	198
microbrew	9	106
glowing	7	87
enfold	6	56
pinfish	7	81

What is the highest possible length of a word?

In [15]:
```
df.describe()
```

Out[15]:

	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

What is the word with the value of `319` ?

In [16]:
```
df.sort_values(by=["Value"],ascending=False)
```

Out[16]:

Word	Char Count	Value
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

In [17]:
```
df.loc[df["Value"]==319]
```

Out[17]:

Word	Char Count	Value
reinstitutionalizations	23	319

What is the most common value?

In [18]:
```
df["Value"].value_counts()
```

Out[18]:
```
Value
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
```

In [19]:
```
df["Value"].mode()
```

Out[19]:
```
0    93
Name: Value, dtype: int64
```

What is the shortest word with value 274 ?

In [21]:
```
df.loc[df["Value"]==274].sort_values(by="Char Count")
```

Out[21]:

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

In [23]:
```
df.loc[(df["Value"]==274)&(df["Char Count"]==20)]

#df.loc[df["Value"]==274,"Char Count"].min()
```

Out[23]:

Word	Char Count	Value
overprotectivenesses	20	274

Create a column `Ratio` which represents the 'Value Ratio' of a word

In [24]:
```
df["Ratio"]=df["Value"]/df["Char Count"]
```

In [25]:
```
df
```

Out[25]:

Word	Char Count	Value	Ratio
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
...	...	...	...
zymotic	7	111	15.857143
zymurgies	9	143	15.888889
zymurgy	7	135	19.285714
zyzzyva	7	151	21.571429
zyzzyvas	8	170	21.250000

172821 rows × 3 columns

What is the maximum value of `Ratio` ?

In [27]:
```
df["Ratio"].max()
```

Out[27]:
```
22.5
```

What word is the one with the highest `Ratio` ?

In [28]:
```
df.loc[df["Ratio"]==22.5]
```

Out[28]:

Word	Char Count	Value	Ratio
xu	2	45	22.5

In [29]:
```
df.sort_values(by="Ratio",ascending=False)
```

Out[29]:

Word	Char Count	Value	Ratio
xu	2	45	22.500000
muzzy	5	111	22.200000
wry	3	66	22.000000
xyst	4	88	22.000000
pyx	3	65	21.666667
...	...	...	...
ab	2	3	1.500000
baba	4	6	1.500000
aba	3	4	1.333333
baa	3	4	1.333333
aa	2	2	1.000000

172821 rows × 3 columns

In [30]:
```
df.loc[df["Ratio"]==df["Ratio"].max()]
```

Out[30]:

Word	Char Count	Value	Ratio
xu	2	45	22.5

How many words have a `Ratio` of 10 ?

In [35]:
```
df[df["Ratio"]==10].count()
```

Out[35]:
```
Char Count    2604
Value         2604
Ratio         2604
dtype: int64
```

In [36]:
```
df.loc[df["Ratio"]==10]
```

Out[36]:

Word	Char Count	Value	Ratio
aardwolf	8	80	10.0
abatements	10	100	10.0
abducts	7	70	10.0
abetment	8	80	10.0
abettals	8	80	10.0
...	...	...	...
ycleped	7	70	10.0
yodeled	7	70	10.0
zamia	5	50	10.0
zebecs	6	60	10.0
zwieback	8	80	10.0

2604 rows × 3 columns

In [38]:
```
df.loc[df["Ratio"]==10].shape
```

Out[38]:
```
(2604, 3)
```

In [37]:
```
df["Ratio"].value_counts()
```

Out[37]:
```
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
```

In [39]:
```
df.query("Ratio==10").shape
```

Out[39]:
```
(2604, 3)
```

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

In [44]:
```
df.query("Ratio==10").sort_values(by="Value",ascending=False)
```

Out[44]:

Word	Char Count	Value	Ratio
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 [45]:
```
df.loc[df["Ratio"]==10,"Value"].max()
```

Out[45]:
```
240
```

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

In [48]:
```
df.query("Value==260").sort_values(by="Char Count")
```

Out[48]:

Word	Char Count	Value	Ratio
hydroxytryptamine	17	260	15.294118
neuropsychologists	18	260	14.444444
psychophysiologist	18	260	14.444444
revolutionarinesses	19	260	13.684211
countermobilizations	20	260	13.000000
underrepresentations	20	260	13.000000

In [49]:
```
df.loc[df["Value"]==260,"Char Count"].min()
```

Out[49]:
```
17
```

Based on the previous task, what word is it?

In [54]:
```
df.loc[(df["Value"]==260)&(df["Char Count"]==17)]
```

Out[54]:

Word	Char Count	Value	Ratio
hydroxytryptamine	17	260	15.294118

Find all the words with Char Count> Avg Char Count

In [55]:
```
mean_char_count=df["Char Count"].mean()
mean_char_count
```

Out[55]:
```
9.087628239623657
```

In [56]:
```
df.query("`Char Count`>9")
```

Out[56]:

Word	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

In [57]:
```
df.query("`Char Count` > @mean_char_count")
```

Out[57]:

Word	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