

**Name : Aditya Prashant Bamnote**

**Roll no. : ET2-79**

**PRN : 202401070204**

**Allotted Dataset : wordnet**

**Theory Activity No. 1**

**Formulated 20 Problem Statements for “wordnet synonyms.csv”**

**1. Show first 5 rows of dataset**

Code : `print("\n1. First 5 rows:\n", df.head())`

Output :

```
1. First 5 rows:
   lemma part_of_speech synonyms
0  .22-caliber    adjective .22 caliber;.22 calibre;.22-calibre
1  .22-calibre    adjective .22 caliber;.22-caliber;.22 calibre
2  .22 caliber    adjective .22-caliber;.22 calibre;.22-calibre
3  .22 calibre    adjective .22 caliber;.22-caliber;.22-calibre
4  .38-caliber    adjective .38 caliber;.38 calibre;.38-calibre
```

**2. Find total number of rows and columns**

Code : `print("\n2. Shape of dataset (rows, columns):", df.shape)`

Output : `2. Shape of dataset (rows, columns): (127001, 3)`

**3. List all unique Parts of Speech**

Code : `print("\n3. Unique Parts of Speech:\n", df['part of speech'].unique())`

Output : `3. Unique Parts of Speech:
['adjective' 'noun' 'satellite' 'adverb' 'verb']`

**4. Count how many words are there for each Part of Speech**

Code :

`print("\n4. Word count for each Part of Speech:\n", df['part_of_speech'].value_counts())`

```

4. Word count for each Part of Speech:
part_of_speech
noun          94600
verb          13681
satellite     11752
adjective     4274
adverb        2694
Name: count, dtype: int64

```

5. Find the lemma which has maximum synonyms

```
Code : df['synonym_count'] = df['synonyms'].apply(lambda x: len(str(x).split(';')))
```

```
max_synonyms_row = df[df['synonym_count'] == df['synonym_count'].max()]
```

```
print("\n5. Lemma with maximum synonyms:\n", max_synonyms_row[['lemma', 'synonyms', 'synonym_count']])
```

Output :

```

5. Lemma with maximum synonyms:
      lemma                                synonyms  synonym_count
13930  passing  pass;go through;go across|travel by;pass by;su...         85
63088  getting  get;acquire|become;go;get|get;let;have|receive...         85

```

6. Find the lemma which has maximum synonyms

Code :

```
min_synonyms_row = df[df['synonym_count'] == df['synonym_count'].min()]
```

```
print("\n6. Lemma with minimum synonyms:\n", min_synonyms_row[['lemma', 'synonyms', 'synonym_count']])
```

Output :

```

6. Lemma with minimum synonyms:
      lemma                                synonyms  synonym_count
13         0                                zero             1
18  10-membered          ten-membered             1
23   1000th        thousandth             1
26   101st  hundred-and-first             1
28   105th  hundred-and-fifth             1
...         ...                                ...             ...
126967  work at          work on             1
126977  wreathe          wind             1
126978  wring from        extort             1
126979  wring out    squeeze out             1
126992  yarn-dye        impress             1
[39559 rows x 3 columns]

```

7. Find average number of synonyms per lemma

Code :

```
print("\n7. Average number of synonyms per lemma:", df['synonym_count'].mean())
```

Output :

```
7. Average number of synonyms per lemma: 2.9391973291548887
```

8. How many lemmas have more than 3 synonyms

Code :

```
print("\n8. Lemmas with more than 3 synonyms:\n", df[df['synonym_count'] > 3])
```

Output :

```
8. Lemmas with more than 3 synonyms:
   lemma part_of_speech synonyms synonym_count
12      0          noun  zero;nought;cipher;cypher      4
14      1          noun    one;I;ace;single;unity      5
16     10          noun    ten;X;tenner;decade      4
19    100          noun  hundred;C;century;one C      4
21   1000          noun thousand;one thousand;M;K;chiliad;G;grand;thou...      9
...     ...          ...          ...          ...
126989 yack away      verb  yack;jaw;rattle on;yap away      4
126990 yammer      verb  howl;wrawl;yowl|whine;grizzle;yawp      5
126991 yap away      verb  yack;jaw;yack away;rattle on      4
126993 yaup      verb  howl;ululate;wail;roar;yawl      5
126995 yearn      verb  hanker;long|ache;yen;pine;languish      5

[29181 rows x 4 columns]
```

9. Sort the lemmas based on number of synonyms (descending)

Code :

```
print("\n9. Lemmas sorted by number of synonyms:\n", df.sort_values(by='synonym_count',
ascending=False))
```

output :

```
9. Lemmas sorted by number of synonyms:
   lemma part_of_speech synonyms synonym_count
63088  getting      verb  get;acquire|become;go;get|get;let;have|receive...      85
13930  passing      verb  pass;go through;go across|travel by;pass by;su...      85
35416  breaking      verb  interrupt;break|break;separate;split up;fall a...      84
3041   broke      verb  interrupt;break|break;separate;split up;fall a...      84
3043   broken      verb  interrupt;break|break;separate;split up;fall a...      84
...     ...          ...          ...          ...
87405  oriel window      noun          oriel      1
87404  oriel      noun          oriel window      1
87432  origen      noun          Origen      1
87427  orientation course      noun          orientation      1
87423  oriental spruce      noun  Picea orientalis      1

[127001 rows x 4 columns]
```

10. Display lemmas where synonym list is missing or empty

Code :

```
print("\n10. Lemmas with missing or empty synonym list:\n", df[df['synonyms'].isnull() |  
(df['synonyms'] == "")])
```

output :

```
10. Lemmas with missing or empty synonym list:  
   lemma part_of_speech synonyms synonym_count  
21394  void      satellite    NaN            1  
83723  naan       noun        NaN            1  
  
11. Filled missing synonym lists:  
   lemma part_of_speech synonyms synonym_count  
0      .22-caliber    adjective .22 caliber;.22 calibre;.22-calibre    3  
1      .22-calibre    adjective .22 caliber;.22-caliber;.22 calibre    3  
2      .22 caliber    adjective .22-caliber;.22 calibre;.22-calibre    3  
3      .22 calibre    adjective .22 caliber;.22-caliber;.22-calibre    3  
4      .38-caliber    adjective .38 caliber;.38 calibre;.38-calibre    3  
...      ...      ...      ...      ...  
126996  zero in      verb      range in;home in|zero    2  
126997  zip by      verb      fly by;whisk by    2  
126998  zip up      verb      zipper;zip    2  
126999  zonk out     verb      pass out;black out    2  
127000  zoom along   verb      zoom;whizz;whizz along    3  
  
[127001 rows x 4 columns]
```

11. Fill missing synonym lists with 'No synonyms available'

Code :

```
df['synonyms'] = df['synonyms'].fillna('No synonyms available')  
  
print("\n11. Filled missing synonym lists:\n", df)
```

12. Convert all lemma names to uppercase

Code :

```
df['lemma_upper'] = df['lemma'].str.upper()  
  
print("\n12. Lemmas in uppercase:\n", df[['lemma', 'lemma_upper']])
```

output :

13. Find lemmas where Part of Speech is 'noun'

Code :

```
print("\n13. Lemmas with Part of Speech 'noun':\n", df[df['part_of_speech'].str.lower() == 'noun'])
```

output :

```
13. Lemmas with Part of Speech 'noun':
      lemma part_of_speech synonyms synonym_count lemma_upper
12      0      noun      zero;nought;cipher;cypher      4      0
14      1      noun      one;I;ace;single;unity      5      1
16     10      noun      ten;X;tenner;decade      4     10
19    100      noun      hundred;C;century;one C      4    100
21   1000      noun      thousand;one thousand;M;K;chiliad;G;grand;thou...      9   1000
...     ...      ...      ...      ...      ...
124758 obsess      noun      Ob;Ob River|obstetrics;OB;tocology;midwifery      5  OBSESS
125129 possess      noun      posse;posse comitatus      2  POSSESS
125730 saw logs      noun      saw log      1  SAW LOGS
125946 siss      noun      sister;sis      2    SISS
126280 succuss      noun      juice;succus      2  SUCCUSS

[94600 rows x 5 columns]
```

14. Replace 'noun' with 'NOUN' for better format

Code :

```
df['part_of_speech'] = df['part_of_speech'].str.replace('noun', 'NOUN')
```

```
print("\n14. Parts of Speech after replacing 'noun' with 'NOUN':\n", df['part_of_speech'])
```

```
14. Parts of Speech after replacing 'noun' with 'NOUN':
0      adjective
1      adjective
2      adjective
3      adjective
4      adjective
...
126996      verb
126997      verb
126998      verb
126999      verb
127000      verb
Name: part_of_speech, Length: 127001, dtype: object
```

output :

15. Find duplicate lemma names (if any)

Code :

```
print("\n15. Duplicate lemmas:\n", df[df.duplicated('lemma')])
```

output :

```

15. Duplicate lemmas:
      lemma part_of_speech      synonyms  synonym_count  lemma_upper
13      0      satellite      zero      1      0
15      1      satellite      one;i;ane      3      1
17      10     satellite      ten;x      2      10
20      100    satellite      hundred;one hundred;c      3      100
22      1000   satellite      thousand;one thousand;m;k      4      1000
...      ...      ...      ...      ...      ...
124759  obsess  verb      haunt;ghost      2      OBSESS
125130  possess  verb      own;have      2      POSSESS
125731  saw logs  verb      snore;saw wood      2      SAW LOGS
125947  siss     verb      hiss;sizz;sibilate|hiss;sizz;sibilate      5      SISS
126281  succuss  verb      shake up      1      SUCCUSS

[9797 rows x 5 columns]

```

## 16. Remove duplicate lemma rows (if any)

Code :

```
df = df.drop_duplicates(subset='lemma')
```

```
print("\n16. Dataset after removing duplicate lemmas:\n", df)
```

output :

```

16. Dataset after removing duplicate lemmas:
      lemma part_of_speech      synonyms  synonym_count  lemma_upper
0      .22-caliber  adjective  .22 caliber;.22 calibre;.22-calibre      3      .22-CALIBER
1      .22-calibre  adjective  .22 caliber;.22-caliber;.22 calibre      3      .22-CALIBRE
2      .22 caliber  adjective  .22-caliber;.22 calibre;.22-calibre      3      .22 CALIBER
3      .22 calibre  adjective  .22 caliber;.22-caliber;.22-calibre      3      .22 CALIBRE
4      .38-caliber  adjective  .38 caliber;.38 calibre;.38-calibre      3      .38-CALIBER
...      ...      ...      ...      ...      ...
126996  zero in     verb      range in;home in|zero      2      ZERO IN
126997  zip by      verb      fly by;whisk by      2      ZIP BY
126998  zip up      verb      zipper;zip      2      ZIP UP
126999  zonk out    verb      pass out;black out      2      ZONK OUT
127000  zoom along  verb      zoom;whizz;whizz along      3      ZOOM ALONG

[117204 rows x 5 columns]

```

## 17. Create a new column showing number of characters in lemma

Code :

```
df['lemma_length'] = df['lemma'].apply(lambda x: len(str(x)) if pd.notnull(x) else 0)
```

```
print("\n17. Lemma length (number of characters):\n", df[['lemma', 'lemma_length']])
```

Output :

```

17. Lemma length (number of characters):
      lemma  lemma_length
0      .22-caliber      11
1      .22-calibre      11
2      .22 caliber      11
3      .22 calibre      11
4      .38-caliber      11
...      ...      ...
126996      zero in      7
126997      zip by      6
126998      zip up      6
126999      zonk out      8
127000      zoom along      10

[117204 rows x 2 columns]

```

18. Find lemmas having length greater than 6 characters

Code :

```
print("\n18. Lemmas with more than 6 characters:\n", df[df['lemma_length'] > 6])
```

Output :

```

18. Lemmas with more than 6 characters:
      lemma  part_of_speech  synonyms  synonym_count  lemma_upper  lemma_length
0      .22-caliber      adjective  .22 caliber;.22 calibre;.22-calibre      3      .22-CALIBER      11
1      .22-calibre      adjective  .22 caliber;.22-calibre;.22 calibre      3      .22-CALIBRE      11
2      .22 caliber      adjective  .22-caliber;.22 calibre;.22-calibre      3      .22 CALIBER      11
3      .22 calibre      adjective  .22 caliber;.22-calibre;.22-calibre      3      .22 CALIBRE      11
4      .38-caliber      adjective  .38 caliber;.38 calibre;.38-calibre      3      .38-CALIBER      11
...      ...      ...      ...      ...      ...
126991      yap away      verb      yack;jaw;yack away;rattle on      4      YAP AWAY      8
126992      yarn-dye      verb      impress      1      YARN-DYE      8
126996      zero in      verb      range in;home in|zero      2      ZERO IN      7
126999      zonk out      verb      pass out;black out      2      ZONK OUT      8
127000      zoom along      verb      zoom;whizz;whizz along      3      ZOOM ALONG      10

[100799 rows x 6 columns]

```

19. How many total synonyms are there (overall)

Code :

```
total_synonyms = np.sum(df['synonym_count'])

print("\n19. Total synonyms across all lemmas:", total_synonyms)
```

output :

```
19. Total synonyms across all lemmas: 328739
```

## 20. Find and display 5 random lemmas

```
print("\n20. 5 Random lemmas:\n", df.sample(5))
```

Output :

```
20. 5 Random lemmas:
      lemma part_of_speech      synonyms synonym_count lemma_upper lemma_length
113355      toy          NOUN  plaything|miniature|toy dog           1         TOY           3
43072    complexity      NOUN      complexness           1    COMPLEXITY          10
127000  zoom along      verb    zoom;whizz;whizz along           3    ZOOM ALONG          10

[100799 rows x 6 columns]
```

# Complete code :



```

1 import pandas as pd
2 import numpy as np
3
4 df = pd.read_csv('wordnet_synonyms.csv')
5
6 print("Dataset:\n", df)
7
8 print("\n--- Problem Statements and Solutions ---\n")
9
10 print("\n1. First 5 rows:\n", df.head())
11
12 print("\n2. Shape of dataset (rows, columns):", df.shape)
13
14 print("\n3. Unique Parts of Speech:\n", df['part_of_speech'].unique())
15
16 print("\n4. Word count for each Part of Speech:\n", df['part_of_speech'].value_counts())
17
18 df['synonym_count'] = df['synonyms'].apply(lambda x: len(str(x).split(';')))
19 max_synonyms_row = df[df['synonym_count'] == df['synonym_count'].max()]
20 print("\n5. Lemma with maximum synonyms:\n", max_synonyms_row[['lemma', 'synonyms', 'synonym_count']])
21
22
23 min_synonyms_row = df[df['synonym_count'] == df['synonym_count'].min()]
24 print("\n6. Lemma with minimum synonyms:\n", min_synonyms_row[['lemma', 'synonyms', 'synonym_count']])
25
26
27 print("\n7. Average number of synonyms per lemma:", df['synonym_count'].mean())
28
29
30 print("\n8. Lemmas with more than 3 synonyms:\n", df[df['synonym_count'] > 3])
31
32
33 print("\n9. Lemmas sorted by number of synonyms:\n", df.sort_values(by='synonym_count', ascending=False))
34
35 print("\n10. Lemmas with missing or empty synonym list:\n", df[df['synonyms'].isnull() | (df['synonyms'] == '')])
36
37 df['synonyms'] = df['synonyms'].fillna('No synonyms available')
38 print("\n11. Filled missing synonym lists:\n", df)
39
40 df['lemma_upper'] = df['lemma'].str.upper()
41 print("\n12. Lemmas in uppercase:\n", df[['lemma', 'lemma_upper']])
42
43 print("\n13. Lemmas with Part of Speech 'noun':\n", df[df['part_of_speech'].str.lower() == 'noun'])
44
45 df['part_of_speech'] = df['part_of_speech'].str.replace('noun', 'NOUN')
46 print("\n14. Parts of Speech after replacing 'noun' with 'NOUN':\n", df['part_of_speech'])
47
48 print("\n15. Duplicate lemmas:\n", df[df.duplicated('lemma')])
49
50 df = df.drop_duplicates(subset='lemma')
51 print("\n16. Dataset after removing duplicate lemmas:\n", df)
52
53 df['lemma_length'] = df['lemma'].apply(lambda x: len(str(x)) if pd.notnull(x) else 0)
54 print("\n17. Lemma length (number of characters):\n", df[['lemma', 'lemma_length']])
55
56 print("\n18. Lemmas with more than 6 characters:\n", df[df['lemma_length'] > 6])
57
58 total_synonyms = np.sum(df['synonym_count'])
59 print("\n19. Total synonyms across all lemmas:", total_synonyms)
60
61 print("\n20. 5 Random lemmas:\n", df.sample(5))
62
63

```

```
PS D:\aditya_bm> py
> python -u "d:\aditya_bm\Theory_activity_1.py"
```

Dataset:

	lemma	part_of_speech	synonyms
0	.22-caliber	adjective	.22 caliber;.22 calibre;.22-calibre
1	.22-calibre	adjective	.22 caliber;.22-caliber;.22 calibre
2	.22 caliber	adjective	.22-caliber;.22 calibre;.22-calibre
3	.22 calibre	adjective	.22 caliber;.22-caliber;.22-calibre
4	.38-caliber	adjective	.38 caliber;.38 calibre;.38-calibre
...	...	...	...
126996	zero in	verb	range in;home in zero
126997	zip by	verb	fly by;whisk by
126998	zip up	verb	zipper;zip
126999	zonk out	verb	pass out;black out
127000	zoom along	verb	zoom;whizz;whizz along

[127001 rows x 3 columns]

--- Problem Statements and Solutions ---

1. First 5 rows:

	lemma	part_of_speech	synonyms
0	.22-caliber	adjective	.22 caliber;.22 calibre;.22-calibre
1	.22-calibre	adjective	.22 caliber;.22-caliber;.22 calibre
2	.22 caliber	adjective	.22-caliber;.22 calibre;.22-calibre
3	.22 calibre	adjective	.22 caliber;.22-caliber;.22-calibre
4	.38-caliber	adjective	.38 caliber;.38 calibre;.38-calibre

2. Shape of dataset (rows, columns): (127001, 3)

3. Unique Parts of Speech:

['adjective' 'noun' 'satellite' 'adverb' 'verb']

4. Word count for each Part of Speech:

part_of_speech	
noun	94600
verb	13681
satellite	11752
adjective	4274
adverb	2694

Name: count, dtype: int64

5. Lemma with maximum synonyms:

	lemma	synonyms	synonym_count
13930	passing	pass;go through;go across travel by;pass by;su...	85
63088	getting	get;acquire become;go;get get;let;have receive...	85

#### 6. Lemma with minimum synonyms:

	lemma	synonyms	synonym_count
13	0	zero	1
18	10-membered	ten-membered	1
23	1000th	thousandth	1
26	101st	hundred-and-first	1
28	105th	hundred-and-fifth	1
...	...	...	...
126967	work at	work on	1
126977	wreathe	wind	1
126978	wring from	extort	1
126979	wring out	squeeze out	1
126992	yarn-dye	impress	1

[39559 rows x 3 columns]

7. Average number of synonyms per lemma: 2.9391973291548887

#### 8. Lemmas with more than 3 synonyms:

	lemma	part_of_speech	synonyms	synonym_count
12	0	noun	zero;nought;cipher;cypher	4
14	1	noun	one;I;ace;single;unity	5
16	10	noun	ten;X;tenner;decade	4
19	100	noun	hundred;C;century;one C	4
21	1000	noun	thousand;one thousand;M;K;chiliad;G;grand;thou...	9
...	...	...	...	...
126989	yack away	verb	yack;jaw;rattle on;yap away	4
126990	yammer	verb	howl;wrawl;yowl whine;grizzle;yawp	5
126991	yap away	verb	yack;jaw;yack away;rattle on	4
126993	yaup	verb	howl;ululate;wail;roar;yawl	5
126995	yearn	verb	hanker;long ache;yen;pine;languish	5

[29181 rows x 4 columns]

#### 9. Lemmas sorted by number of synonyms:

	lemma	part_of_speech	synonyms	synonym_count
63088	getting	verb	get;acquire become;go;get get;let;have receive...	85
13930	passing	verb	pass;go through;go across travel by;pass by;su...	85
35416	breaking	verb	interrupt;break break;separate;split up;fall a...	84
3041	broke	verb	interrupt;break break;separate;split up;fall a...	84
3043	broken	verb	interrupt;break break;separate;split up;fall a...	84
...	...	...	...	...
87405	oriel window	noun	oriel	1
87404	oriel	noun	oriel window	1
87432	origen	noun	Origen	1
87427	orientation course	noun	orientation	1
87423	oriental spruce	noun	Picea orientalis	1

[127001 rows x 4 columns]

#### 10. Lemmas with missing or empty synonym list:

	lemma	part_of_speech	synonyms	synonym_count
21394	void	satellite	NaN	1
83723	naan	noun	NaN	1

#### 10. Lemmas with missing or empty synonym list:

	lemma	part_of_speech	synonyms	synonym_count
21394	void	satellite	NaN	1
83723	naan	noun	NaN	1

#### 11. Filled missing synonym lists:

	lemma	part_of_speech	synonyms	synonym_count
0	.22-caliber	adjective	.22 caliber;.22 calibre;.22-calibre	3
1	.22-calibre	adjective	.22 caliber;.22-caliber;.22 calibre	3
2	.22 caliber	adjective	.22-caliber;.22 calibre;.22-calibre	3
3	.22 calibre	adjective	.22 caliber;.22-caliber;.22-calibre	3
4	.38-caliber	adjective	.38 caliber;.38 calibre;.38-calibre	3
...	...	...	...	...
126996	zero in	verb	range in;home in zero	2
126997	zip by	verb	fly by;whisk by	2
126998	zip up	verb	zipper;zip	2
126999	zonk out	verb	pass out;black out	2
127000	zoom along	verb	zoom;whizz;whizz along	3

[127001 rows x 4 columns]

#### 12. Lemmas in uppercase:

	lemma	lemma_upper
0	.22-caliber	.22-CALIBER
1	.22-calibre	.22-CALIBRE
2	.22 caliber	.22 CALIBER
3	.22 calibre	.22 CALIBRE
4	.38-caliber	.38-CALIBER
...	...	...
126996	zero in	ZERO IN
126997	zip by	ZIP BY
126998	zip up	ZIP UP
126999	zonk out	ZONK OUT
127000	zoom along	ZOOM ALONG

[127001 rows x 2 columns]

#### 13. Lemmas with Part of Speech 'noun':

	lemma	part_of_speech	synonyms	synonym_count	lemma_upper
12	0	noun	zero;nought;cipher;cypher	4	0
14	1	noun	one;I;ace;single;unity	5	1
16	10	noun	ten;X;tenner;decade	4	10
19	100	noun	hundred;C;century;one C	4	100
21	1000	noun	thousand;one thousand;M;K;chiliad;G;grand;thou...	9	1000
...	...	...	...	...	...
124758	obsess	noun	Ob;Ob River obstetrics;OB;tocology;midwifery	5	OBSESS
125129	possess	noun	posse;posse comitatus	2	POSSESS
125730	saw logs	noun	saw log	1	SAW LOGS
125946	siss	noun	sister;sis	2	SISS
126280	succuss	noun	juice;succus	2	SUCCUSS

[94600 rows x 5 columns]

14. Parts of Speech after replacing 'noun' with 'NOUN':

```
0      adjective
1      adjective
2      adjective
3      adjective
4      adjective
```

...

```
126996      verb
126997      verb
126998      verb
126999      verb
127000      verb
```

Name: part\_of\_speech, Length: 127001, dtype: object

15. Duplicate lemmas:

	lemma	part_of_speech	synonyms	synonym_count	lemma_upper
13	0	satellite	zero	1	0
15	1	satellite	one;i;ane	3	1
17	10	satellite	ten;x	2	10
20	100	satellite	hundred;one hundred;c	3	100
22	1000	satellite	thousand;one thousand;m;k	4	1000
...	...	...	...	...	...
124759	obsess	verb	haunt;ghost	2	OBSESS
125130	possess	verb	own;have	2	POSSESS
125731	saw logs	verb	snore;saw wood	2	SAW LOGS
125947	sisss	verb	hiss;sizz;sibilate hiss;sizz;sibilate	5	SISS
126281	succuss	verb	shake up	1	SUCCUSS

[9797 rows x 5 columns]

16. Dataset after removing duplicate lemmas:

	lemma	part_of_speech	synonyms	synonym_count	lemma_upper
0	.22-caliber	adjective	.22 caliber;.22 calibre;.22-calibre	3	.22-CALIBER
1	.22-calibre	adjective	.22 caliber;.22-caliber;.22 calibre	3	.22-CALIBRE
2	.22 caliber	adjective	.22-caliber;.22 calibre;.22-calibre	3	.22 CALIBER
3	.22 calibre	adjective	.22 caliber;.22-caliber;.22-calibre	3	.22 CALIBRE
4	.38-caliber	adjective	.38 caliber;.38 calibre;.38-calibre	3	.38-CALIBER
...	...	...	...	...	...
126996	zero in	verb	range in;home in zero	2	ZERO IN
126997	zip by	verb	fly by;whisk by	2	ZIP BY
126998	zip up	verb	zipper;zip	2	ZIP UP
126999	zonk out	verb	pass out;black out	2	ZONK OUT
127000	zoom along	verb	zoom;whizz;whizz along	3	ZOOM ALONG

[117204 rows x 5 columns]

17. Lemma length (number of characters):

	lemma	lemma_length
0	.22-caliber	11
1	.22-calibre	11
2	.22 caliber	11
3	.22 calibre	11
4	.38-caliber	11
...	...	...
126996	zero in	7
126997	zip by	6
126998	zip up	6
126999	zonk out	8
127000	zoom along	10

[117204 rows x 2 columns]

18. Lemmas with more than 6 characters:

	lemma	part_of_speech	synonyms	synonym_count	lemma_upper	lemma_length
0	.22-caliber	adjective	.22 caliber;.22 calibre;.22-calibre	3	.22-CALIBER	11
1	.22-calibre	adjective	.22 caliber;.22-caliber;.22 calibre	3	.22-CALIBRE	11
2	.22 caliber	adjective	.22-caliber;.22 calibre;.22-calibre	3	.22 CALIBER	11
3	.22 calibre	adjective	.22 caliber;.22-caliber;.22-calibre	3	.22 CALIBRE	11
4	.38-caliber	adjective	.38 caliber;.38 calibre;.38-calibre	3	.38-CALIBER	11
...	...	...	...	...	...	...
126991	yap away	verb	yack;jaw;yack away;rattle on	4	YAP AWAY	8
126992	yarn-dye	verb	impress	1	YARN-DYE	8
126996	zero in	verb	range in;home in zero	2	ZERO IN	7
126999	zonk out	verb	pass out;black out	2	ZONK OUT	8
127000	zoom along	verb	zoom;whizz;whizz along	3	ZOOM ALONG	10

[100799 rows x 6 columns]

19. Total synonyms across all lemmas: 328739

20. 5 Random lemmas:

	lemma	part_of_speech	synonyms	synonym_count	lemma_upper	lemma_length
3145	bullet-headed	satellite	bullheaded;pigheaded	2	BULLET-HEADED	13
34049	blood	NOUN	rake;rakehell;profligate;rip;roue lineage;line...	16	BLOOD	5
74726	lake malawi	NOUN	Lake Nyasa;Lake Malawi	2	LAKE MALAWI	11
51817	enlisted officer	NOUN	noncommissioned officer;noncom	2	ENLISTED OFFICER	16
84286	negativist	NOUN	defeatist	1	NEGATIVIST	10

PS D:\aditya\_bm>