Name : Shreyash Sainath Telang

Class : CS5-80

PRN : 202401100006

JupyterLab ⬚ ⚙ Python 3 (ipykernel) ○ ≡

```
Name : Shreyash Sainath Telang
Class : CS5-80
PRN : 202401100006
Topic : Formulate 20 problem statements for a wordnet(Wordned Synonyms) dataset using Numpy and Pandas and Apply Numpy and pandas methods to find the
solution for the formulated problem statements.
```

1. Importing pandas and numpy and reading dataset

```python
[4]: import pandas as pd
     import numpy as np

     df = pd.read_csv('WordnetSynonyms.csv')
     print(df)
```

```
              Word  Count       POS  \
0             a-ok      4  satellite
1           a-okay      6  satellite
2            a-one      5  satellite
3             a.m.      4  satellite
4             a.m.      4     adverb
...            ...    ...       ...
126722     zero in      7      verb
126723      zip by      6      verb
126724      zip up      6      verb
126725    zonk out      8      verb
126726  zoom along     10      verb

                                           Synonyms
0                                            a-okay
1                                              a-ok
2         ace;A-one;crack;first-rate;super;tiptop;topnot...
3                                      ante meridiem
4                            ante meridiem;A.M.
...                                             ...
126722                      range in;home in|zero
126723                         fly by;whisk by
126724                               zipper;zip
126725                        pass out;black out
126726                    zoom;whizz;whizz along

[126727 rows x 4 columns]
```

```
[ ]:
```

1. Total entries

```python
[7]: print("\n1. Total entries in the dataset:")
     print(len(df))
```

```
1. Total entries in the dataset:
126727
```

2. Unique words

```python
[9]: print("\n2. Total unique words:")
     print(df['Word'].nunique())
```

```
2. Total unique words:
116966
```

3. Missing values in each column

```python
[10]: print("\n3. Missing values in each column:")
      print(df.isnull().sum())
```

```
3. Missing values in each column:
Word         3
Count        0
POS          0
Synonyms     2
dtype: int64
```

4. Unique POS categories

```
[11]:  print("\n4. Unique POS categories:")
       print(df['POS'].unique())
```

```
4. Unique POS categories:
['satellite' 'adverb' 'verb' 'adjective' 'noun']
```

5. Word with the maximum synonym count

```
[13]:  df['Synonym_Count'] = df['Synonyms'].apply(lambda x: len(str(x).split(';')) if pd.notnull(x) else 0)
       print("\n5. Word with the maximum number of synonyms:")
       print(df.loc[df['Synonym_Count'].idxmax()])
```

```
5. Word with the maximum number of synonyms:
Word                                          passing
Count                                               7
POS                                              verb
Synonyms          pass;go through;go across|travel by;pass by;su...
Synonym_Count                                      85
Name: 13656, dtype: object
```

6. Average synonym count

```
[14]:  print("\n6. Average synonym count across entries:")
       print(df['Synonym_Count'].mean())
```

```
6. Average synonym count across entries:
2.9411333022954858
```

7. Total number of different synonyms

```
[15]:  print("\n7. Total number of different synonyms listed:")
       all_synonyms = set(';'.join(df['Synonyms'].dropna()).split(';'))
       print(len(all_synonyms))
```

```
7. Total number of different synonyms listed:
139298
```

8. Most common POS tag

```
[16]:  print("\n8. Most common POS tag:")
       print(df['POS'].mode()[0])
```

```
8. Most common POS tag:
noun
```

9. Average frequency (Count) of the words

```
[17]:  print("\n9. Average frequency (Count) of the words:")
       print(df['Count'].mean())
```

```
9. Average frequency (Count) of the words:
11.253237273824835
```

10. 5 words with the highest count

```
[18]:  print("\n10. Top 5 words with highest 'Count':")
       print(df.sort_values(by='Count', ascending=False).head(5)[['Word', 'Count']])
```

```
10. Top 5 words with highest 'Count':
                                        Word  Count
33776    blood-oxygenation level dependent functional m...     71
27086    american federation of labor and congress of i...     69
70734    international islamic front for jihad against ...     64
83730    national association of securities dealers aut...     63
115120   united nations educational scientific and cult...     63
```

```
# 11. Words with exactly 1 synonym
```

```
[19]:  print("\n11. Words with exactly 1 synonym:")
       print(df[df['Synonym_Count'] == 1][['Word', 'Synonyms']])
```

```
11. Words with exactly 1 synonym:
                 Word      Synonyms
0                a-ok       a-okay
1              a-okay         a-ok
3                a.m.  ante meridiem
5          a couple of       a few
6               a few  a couple of
...               ...          ...
126693         work at      work on
126703         wreathe         wind
126704       wring from       extort
126705        wring out   squeeze out
126718         yarn-dye       impress

[39467 rows x 2 columns]
```

```
# 12. Number of 'satellite' POS words
```

```
[20]:  print("\n12. Number of 'satellite' POS words:")
       print(df[df['POS'] == 'satellite'].shape[0])
```

```
12. Number of 'satellite' POS words:
11532
```

```
# 13. All adjectives (adjective / adjective satellite)
```

```
[21]:  print("\n13. List of all adjectives:")
       print(df[df['POS'].str.contains('adjective', case=False, na=False)][['Word', 'POS']])
```

```
13. List of all adjectives:
              Word        POS
12           abasic   adjective
13           abatic   adjective
14           abaxial  adjective
20          abducent  adjective
22         abducting  adjective
...             ...        ...
118272    westerner  adjective
118290       wetter  adjective
118906       wilder  adjective
123088       flatter  adjective
126489      unsolder  adjective

[4262 rows x 2 columns]
```

```
# 14. Correlation between Count and Synonym_Count
```

```
[22]:  print("\n14. Correlation between 'Count' and 'Synonym_Count':")
       print(df['Count'].corr(df['Synonym_Count']))
```

```
14. Correlation between 'Count' and 'Synonym_Count':
-0.15077623330814147
```

```
# 15. Words with no synonyms listed
```

```
[23]:  print("\n15. Number of words with no synonyms listed:")
       print(df['Synonyms'].isnull().sum())
```

```
15. Number of words with no synonyms listed:
2
```

```python
[24]: print("\n16. Duplicate entries based on 'Word':")
      print(df[df.duplicated(subset=['Word'], keep=False)].sort_values('Word'))
```

```
16. Duplicate entries based on 'Word':
               Word  Count        POS  \
54978        02-Feb      5       noun
54971        02-Feb      5       noun
85346        11-Nov      5       noun
24362        11-Nov      5       noun
102462       11-Sep      5       noun
...             ...    ...        ...
21875    zoroastrian     11  adjective
21874    zoroastrian     11       noun
12828           NaN      4       noun
12829           NaN      4   satellite
83557           NaN      3       noun

                                            Synonyms  Synonym_Count
54978                       Groundhog Day;February 2              2
54971                  Candlemas;Candlemas Day;Feb 2              3
85346    Veterans Day;Veterans' Day;Armistice Day;Novem...              4
24362            Martinmas;St Martin's Day;11 November              3
102462           9/11;9-11;September 11;Sept. 11;Sep 11              5
...                                              ...            ...
21875                                    Zoroastrian              1
21874                                    Zoroastrian              1
12828    nothing;nil;nix;nada;aught;cipher;cypher;goose...             13
12829                                           void              1
83557    grandma;grandmother;granny;grannie;gran;nanna|...              7

[18674 rows x 5 columns]
```

```python
[25]: print("\n17. Synonyms for the word 'happy':")
      print(df[df['Word'] == 'happy']['Synonyms'].tolist())
```

```
17. Synonyms for the word 'happy':
['felicitous|glad|well-chosen']
```

18. Words whose synonyms include 'joy'

```python
print("\n18. Words whose synonyms include 'joy':")
print(df[df['Synonyms'].str.contains('joy', na=False)][['Word', 'Synonyms']])
```

```
18. Words whose synonyms include 'joy':
                    Word                                           Synonyms
5152           delighted  please;delight|delight;enjoy;revel|enchant;enr...
6294               elated                          gleeful;joyful;jubilant
6983             exulting  exult;walk on air;be on cloud nine;jump for jo...
8165            gladdened                             gladden;joy|gladden
8184              gleeful                          elated;joyful;jubilant
8347            gratifying                     sweet|enjoyable;pleasurable
10377            jubilant  exultant;exulting;prideful;rejoicing;triumphal...
10691            laughing                  laugh;express joy;express mirth
11109                loved  love|love;enjoy|love|sleep together;roll in th...
11116               loving  love|love;enjoy|love|sleep together;roll in th...
13386            overjoyed                                          overjoy
14155          pleasurable                           enjoyable;gratifying
15414            rejoicing  rejoice;joy|exuberate;exult;rejoice;triumph;ju...
21946            agreeably                           pleasantly;enjoyably
22756            gleefully                           joyously;joyfully
23111             joyfully                           gleefully;joyously
23112             joyously                           gleefully;joyfully
23575           pleasantly              cheerily;sunnily|agreeably;enjoyably
41274    clematis vitalba  traveler's joy;traveller's joy;old man's beard...
43489         control stick                               stick;joystick
46862           delectation                             delight|enjoyment
46883               delight                       delectation|joy;pleasure
46884               delight  please|enjoy;revel|enchant;enrapture;transport...
72644                   joy               joyousness;joyfulness|delight;pleasure
72648           joyfulness                              joy;joyousness
72649           joyousness                              joy;joyfulness
74973                laugh                    express joy;express mirth
77389                 love  enjoy|sleep together;roll in the hay;make out;...
86122       old man's beard  Spanish moss;black moss;long moss;Tillandsia u...
88848         party pooper                 spoilsport;killjoy;wet blanket
91889             pleasure                       pleasance|joy;delight
97231               relish                    enjoy;bask;savor;savour
97233            relishing                enjoy;bask;relish;savor;savour
97801                 revel  delight;enjoy|racket;make whoopie;make merry;m...
100820               savor  enjoy;bask;relish;savour|taste;savor|savour|s...
100823             savoring  enjoy;bask;relish;savor;savour|taste;savor;sav...
100826              savour   taste;savor|savor|savor|enjoy;bask;relish;savor
100828            savouring  taste;savor;savour|savor;savour|savor;savour|e...
106556            spoilsport                  killjoy;wet blanket;party pooper
107671                stick  control stick;joystick|pin;peg|joint;marijuana...
112790                 tool                         joyride;tool around
113437        traveler's joy   traveller's joy;old man's beard;Clematis vitalba
113447       traveller's joy    traveler's joy;old man's beard;Clematis vitalba
115555                  use  usage;utilization;utilisation;employment;exerc...
118284          wet blanket                 spoilsport;killjoy;party pooper
121132                 bask                   enjoy;relish;savor;savour
121147       be on cloud nine            exult;walk on air;jump for joy
122927          express mirth                        laugh;express joy
122953                exult  walk on air;be on cloud nine;jump for joy|exub...
123353              gladden                                          joy
125216              rejoice  joy|exuberate;exult;triumph;jubilate|wallow;tr...
126291           tool around                           joyride;tool
126595          walk on air              exult;be on cloud nine;jump for joy
```

19. Average Count per POS

```python
print("\n19. Average 'Count' per POS:")
print(df.groupby('POS')['Count'].mean().sort_values(ascending=False))
```

```
19. Average 'Count' per POS:
POS
noun         12.130587
adverb        9.759094
adjective     9.369779
satellite     9.067464
verb          7.912726
Name: Count, dtype: float64
```

20. Find the number of words that have more than 5 synonyms

```python
print("\n20. Number of words having more than 5 synonyms:")
words_more_than_5_synonyms = df[df['Synonym_Count'] > 5]
print(len(words_more_than_5_synonyms))
```

```
20. Number of words having more than 5 synonyms:
12818
```