```
In [2]: import pandas
        data='train_data.csv'
        names=['text','id']
        dataset=pandas.read_csv(data,names=names)
        print(dataset.shape)

        (41570, 2)
```

```
In [3]: print(dataset.head(20))
```

```
                                                  text     id
0                                                 text     id
1     Keep your gloves, hats, coats and jackets toge...  122885
2     The Home Dynamix Serendipity Ivory 5 ft. 2 in....  188958
3     The Bosch 18-Volt lithium-ion line of Cordless...  146065
4     Restore your Porter-Cable sander or polisher t...  165138
5     The SPIKECUBE Surge Suppressor from Tripp Lite...  185565
6     These speed nuts are made of stamped spring st...  203659
7     Safer Brand 32 oz. Grub Killer Ready to Spray ...  208694
8     7 x 19 wire rope is constructed of seven stran...  164481
9     Barclay's Darla Freestanding Toilet Paper Hold...  173391
10    The Strength rug pad offers up to 10 years of ...  167377
11    With the DEWALT 3/8 in. Pistol Grip Drill Kit,...  156131
12    Farmstead Maple T-Molding is used to transitio...  161735
13    The Aqua Uno makes a statement with a distinct...  109679
14    Raco Bushing is for use on threaded Rigid/IMC ...  184521
15    Cap A Tread is a durable laminate with pre-att...  216746
16    Fusions Hardware's Elite Collection combines t...  174951
17    The Cadet Pressure-Assisted Toilet really is t...  172018
18    No other decorating element can dramatically c...  210777
19    Illuminate ground-level objects and pathways o...  101597
```

```
In [4]: print(dataset.describe())
```

```
                                              text      id
count                                        41570   41570
unique                                       41570   41570
top       This HD Integrated Outdoor IR Speed Dome is bu...  162281
freq                                             1       1
```

```python
In [11]: import pandas

import matplotlib.pyplot as plt
data='train_label.csv'
names=['id','label']
dataset=pandas.read_csv(data,names=names)
print(dataset.shape)
```

```
(75171, 2)
```

```python
In [12]: print(dataset.head(20))
```

```
        id              label
0       id              label
1    100003              Shape
2    100004    Voltage (volts)
3    100004    Wattage (watts)
4    100006    Wattage (watts)
5    100007  ENERGY STAR Certified
6    100008             Finish
7    100008     Indoor/Outdoor
8    100008   Package Quantity
9    100011           Features
10   100011           Included
11   100012           Features
12   100012   Hardware Included
13   100012           Included
14   100015   Package Quantity
15   100016              Shape
16   100017     Indoor/Outdoor
17   100017   Package Quantity
18   100022  ENERGY STAR Certified
19   100022    Voltage (volts)
```

```python
In [13]: import pandas
data='test_data.csv'
names=['text','id']
dataset=pandas.read_csv(data,names=names)
print(dataset.shape)
```

```
(10394, 2)
```

```
In [14]: print(dataset.head(20))
                                                       text      id
         0                                             text      id
         1     These machine screw nuts are designed to be us...  114689
         2     The M18 FUEL Drill/Driver is the Most Powerful...  183172
         3     Steel City 2-Gang 30 cu. in. Square Electrical...  217304
         4     Native Collection Plus has Shaw's SilentStep P...  184115
         5     Fasade decorative 4 ft. x 8 ft. vinyl wall pan...  103786
         6     Makita's 7.2-Volt Lithium-Ion Cordless Hex Dri...  202789
         7     The Penelope bath vanity in premium laminated ...  208062
         8     Everbilt anti-siphon frost proof sillcock assu...  177644
         9     This Professional High Temperature Glue Gun us...  117538
         10    The SureSill 4-9/16 in. x 150 in. White PVC Sl...  198085
         11    Multi-Purpose 2-Tier Wheeled Utility Cart. You...  151405
         12    Complete your mini split by covering the line ...  168677
         13    Keep your outdoor space warm and comfortable w...  138021
         14    The Hilti TE 1500-AVR Performance Package will...  215825
         15    Carlon 1-1/4 in. 90-Degree Sch. 40 Standard El...  137724
         16    Schlage Latitude Matte Black door levers are p...  155792
         17    Charleston outdoor lanterns are specially desi...  169327
         18    The Leland Bath Collection reinterprets the ti...  123993
         19    Blue-green silk shade. Coordinates with Thomas...  199552
```

```
In [15]: import pandas
         import matplotlib.pyplot as plt
         data='train_label.csv'
         names=['id','label']
         dataset=pandas.read_csv(data,names=names)
         print dataset.head(10)
```
```
               id               label
         0      id               label
         1  100003               Shape
         2  100004      Voltage (volts)
         3  100004      Wattage (watts)
         4  100006      Wattage (watts)
         5  100007  ENERGY STAR Certified
         6  100008               Finish
         7  100008        Indoor/Outdoor
         8  100008      Package Quantity
         9  100011             Features
```

```python
import pandas
import matplotlib.pyplot as plt
data='sample_submission.csv'
names=['id']
dataset=pandas.read_csv(data,names=names)
print dataset.head(10)
```

```
                                                                          id
   id      Indoor/Outdoor Commercial / Residential ENERGY STAR Certified Hardware Included Package Quantity Flooring Product Type Co
lor Tools Product Type Included Voltage (volts) Assembly Required Features Wattage (watts) Finish  Shape
114689 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
183172 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
217304 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
184115 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
103786 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
202789 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
208062 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
177644 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
117538 0.0            0.0                          0.0                      0.0               0.0                  0.0               0.
0   0.0            0.0       0.0              0.0                  0.0       0.0              0.0               0.0
```

```
In [23]:  import nltk
          from nltk.tokenize import sent_tokenize
          text=""" These machine screw nuts are designed to be us...  114689
          2    The M18 FUEL Drill/Driver is the Most Powerful...  183172
          3    Steel City 2-Gang 30 cu. in. Square Electrical...  217304
          4    Native Collection Plus has Shaw's SilentStep P...  184115
          5    Fasade decorative 4 ft. x 8 ft. vinyl wall pan...  103786
          6    Makita's 7.2-Volt Lithium-Ion Cordless Hex Dri...  202789
          7    The Penelope bath vanity in premium laminated ...  208062
          8    Everbilt anti-siphon frost proof sillcock assu...  177644
          9    This Professional High Temperature Glue Gun us...  117538
          10   The SureSill 4-9/16 in. x 150 in. White PVC Sl...  198085
          11   Multi-Purpose 2-Tier Wheeled Utility Cart. You...  151405
          12   Complete your mini split by covering the line ...  168677
          13   Keep your outdoor space warm and comfortable w...  138021
          14   The Hilti TE 1500-AVR Performance Package will...  215825
          15   Carlon 1-1/4 in. 90-Degree Sch. 40 Standard El...  137724
          16   Schlage Latitude Matte Black door levers are p...  155792
          17   Charleston outdoor lanterns are specially desi...  169327
          18   The Leland Bath Collection reinterprets the ti...  123993
          19   Blue-green silk shade. Coordinates with Thomas...  199552"""
          tokenized_text=sent_tokenize(text)
          print(tokenized_text)
```

[' These machine screw nuts are designed to be us... 114689\n2   The M18 FUEL Drill/Driver is the Most Powerful... 183172\n3
Steel City 2-Gang 30 cu.', 'in.', "Square Electrical... 217304\n4   Native Collection Plus has Shaw's SilentStep P... 184115\n
5   Fasade decorative 4 ft. x 8 ft. vinyl wall pan... 103786\n6   Makita's 7.2-Volt Lithium-Ion Cordless Hex Dri... 202789\n7
  The Penelope bath vanity in premium laminated ... 208062\n8   Everbilt anti-siphon frost proof sillcock assu... 177644\n9
This Professional High Temperature Glue Gun us... 117538\n10 The SureSill 4-9/16 in.", 'x 150 in.', 'White PVC Sl... 198085\n
11 Multi-Purpose 2-Tier Wheeled Utility Cart.', 'You... 151405\n12 Complete your mini split by covering the line ... 168677
\n13 Keep your outdoor space warm and comfortable w... 138021\n14 The Hilti TE 1500-AVR Performance Package will... 215825\n
15 Carlon 1-1/4 in.', '90-Degree Sch.', '40 Standard El... 137724\n16 Schlage Latitude Matte Black door levers are p... 1557
92\n17 Charleston outdoor lanterns are specially desi... 169327\n18 The Leland Bath Collection reinterprets the ti... 123993
\n19 Blue-green silk shade.', 'Coordinates with Thomas... 199552']

```
In [24]:   from nltk.tokenize import word_tokenize
           tokenized_word=word_tokenize(text)
           print(tokenized_word)
```

```
['These', 'machine', 'screw', 'nuts', 'are', 'designed', 'to', 'be', 'us', '...', '114689', '2', 'The', 'M18', 'FUEL', 'Drill/Dr
iver', 'is', 'the', 'Most', 'Powerful', '...', '183172', '3', 'Steel', 'City', '2-Gang', '30', 'cu', '.', 'in', '.', 'Square',
'Electrical', '...', '217304', '4', 'Native', 'Collection', 'Plus', 'has', 'Shaw', "'s", 'SilentStep', 'P', '...', '184115',
'5', 'Fasade', 'decorative', '4', 'ft.', 'x', '8', 'ft.', 'vinyl', 'wall', 'pan', '...', '103786', '6', 'Makita', "'s", '7.2-Vol
t', 'Lithium-Ion', 'Cordless', 'Hex', 'Dri', '...', '202789', '7', 'The', 'Penelope', 'bath', 'vanity', 'in', 'premium', 'lamina
ted', '...', '208062', '8', 'Everbilt', 'anti-siphon', 'frost', 'proof', 'sillcock', 'assu', '...', '177644', '9', 'This', 'Prof
essional', 'High', 'Temperature', 'Glue', 'Gun', 'us', '...', '117538', '10', 'The', 'SureSill', '4-9/16', 'in', '.', 'x', '15
0', 'in', '.', 'White', 'PVC', 'Sl', '...', '198085', '11', 'Multi-Purpose', '2-Tier', 'Wheeled', 'Utility', 'Cart', '.', 'You',
'...', '151405', '12', 'Complete', 'your', 'mini', 'split', 'by', 'covering', 'the', 'line', '...', '168677', '13', 'Keep', 'you
r', 'outdoor', 'space', 'warm', 'and', 'comfortable', 'w', '...', '138021', '14', 'The', 'Hilti', 'TE', '1500-AVR', 'Performanc
e', 'Package', 'will', '...', '215825', '15', 'Carlon', '1-1/4', 'in', '.', '90-Degree', 'Sch', '.', '40', 'Standard', 'El',
'...', '137724', '16', 'Schlage', 'Latitude', 'Matte', 'Black', 'door', 'levers', 'are', 'p', '...', '155792', '17', 'Charlesto
n', 'outdoor', 'lanterns', 'are', 'specially', 'desi', '...', '169327', '18', 'The', 'Leland', 'Bath', 'Collection', 'reinterpre
ts', 'the', 'ti', '...', '123993', '19', 'Blue-green', 'silk', 'shade', '.', 'Coordinates', 'with', 'Thomas', '...', '199552']
```

```
In [26]:   from nltk.probability import FreqDist
           fdist = FreqDist(tokenized_word)
           print(fdist)
           fdist.most_common(2)
```

```
<FreqDist with 162 samples and 208 outcomes>
```

```
Out[26]:   [('...', 19), ('.', 8)]
```

```
In [27]: import matplotlib.pyplot as plt
         fdist.plot(30,cumulative=False)
         plt.show()
```