# Tutorial Task

Name – Swati Rai

Reg No – 20BCE0996

Date – 13th Jan 2023

## Task 1

```
# Task 1
```

```
In [ ]: #SWATI_20BCE0996
```

```
In [1]: def gender_features(word):
            return {'last_letter': word[-1]}
```

```
In [2]: gender_features('Swati')
Out[2]: {'last_letter': 'i'}
```

```
In [3]: from nltk.corpus import names
```

```
In [6]: labled_names = ([(name, 'male') for name in names.words('male.txt')] + [(name,'female') for name in names.words('female.txt')])
```

```
In [7]: import random
```

```
In [9]: random.shuffle(labled_names)
```

```
In [10]: labled_names
```

```
('Maren', 'female'),
('Porter', 'male'),
('Wilmette', 'female'),
('Wyn', 'male'),
('Tedra', 'female'),
('Vijay', 'male'),
('Tudor', 'male'),
('Bethena', 'female'),
('Smitty', 'male'),
('Randee', 'female'),
('Hiralal', 'male'),
('Giffer', 'male'),
('Haywood', 'male'),
('Katti', 'female'),
('Pail', 'male'),
('Andie', 'female'),
('Glynn', 'male'),
('Hasty', 'male'),
('Michelina', 'female'),
('Helene', 'female')
```

```
In [11]: featuresets = [(gender_features(n), gender) for (n,gender) in labled_names]
```

```
In [12]: featuresets
```

```
Out[12]: [({'last_letter': 'e'}, 'female'),
          ({'last_letter': 'y'}, 'male'),
          ({'last_letter': 'e'}, 'female'),
          ({'last_letter': 'n'}, 'male'),
          ({'last_letter': 'a'}, 'female'),
          ({'last_letter': 'n'}, 'female'),
          ({'last_letter': 'l'}, 'male'),
          ({'last_letter': 'h'}, 'female'),
          ({'last_letter': 'd'}, 'male'),
          ({'last_letter': 'a'}, 'female'),
          ({'last_letter': 'h'}, 'male'),
          ({'last_letter': 'd'}, 'male'),
          ({'last_letter': 'a'}, 'female'),
          ({'last_letter': 'l'}, 'male'),
          ({'last_letter': 'n'}, 'male'),
          ({'last_letter': 'a'}, 'female'),
          ({'last_letter': 's'}, 'female'),
          ({'last_letter': 'n'}, 'female'),
          ({'last_letter': 'a'}, 'female'),
```

```
In [36]: train_set, test_set = featuresets[500:], featuresets[:500]
```

# Task 2

```
# Task 2

In [37]: import nltk

In [41]: classifier = nltk.NaiveBayesClassifier.train(train_set)

In [47]: classifier.classify(gender_features('Swati'))
Out[47]: 'female'

In [ ]: print(nltk.classify.accuracy(classifier, test_set))

In [44]: import nltk
         from nltk.tokenize import TweetTokenizer
         text = 'The party was sooo fun :D #superfun'
         twtkn = TweetTokenizer()
         twtkn.tokenize(text)
Out[44]: ['The', 'party', 'was', 'sooo', 'fun', ':D', '#superfun']

In [ ]:
```

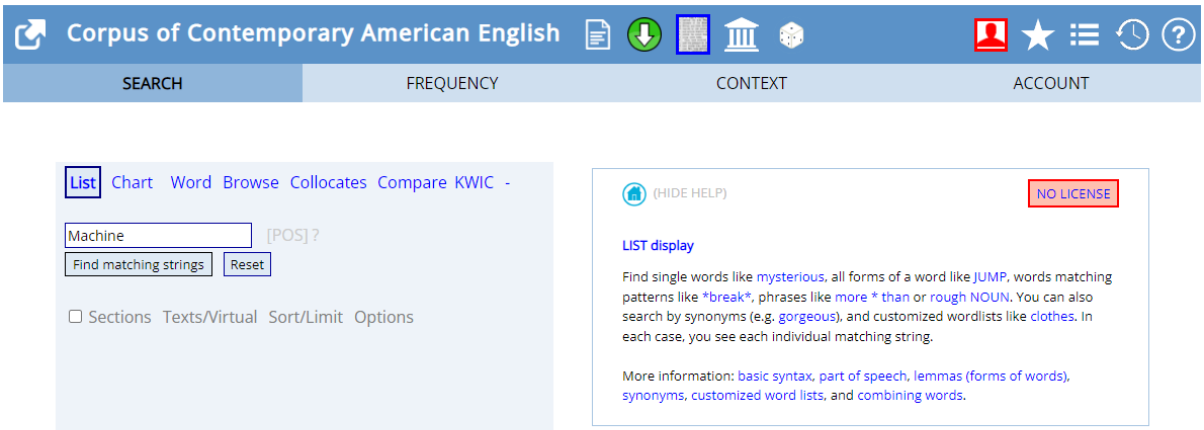
# Task 3

## Explore COCA

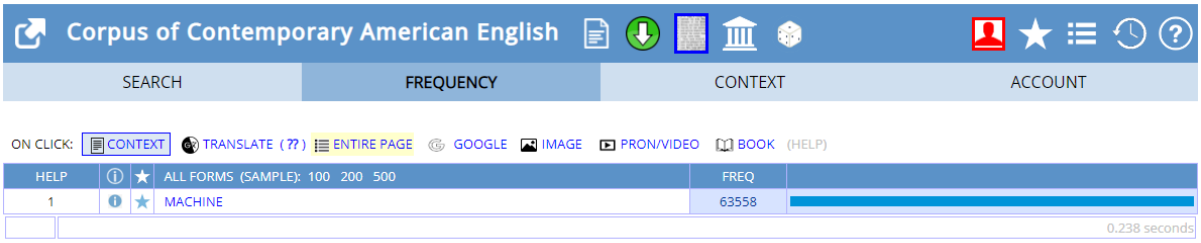
Demonstrate the following features

3.1 Find the frequency count of a word

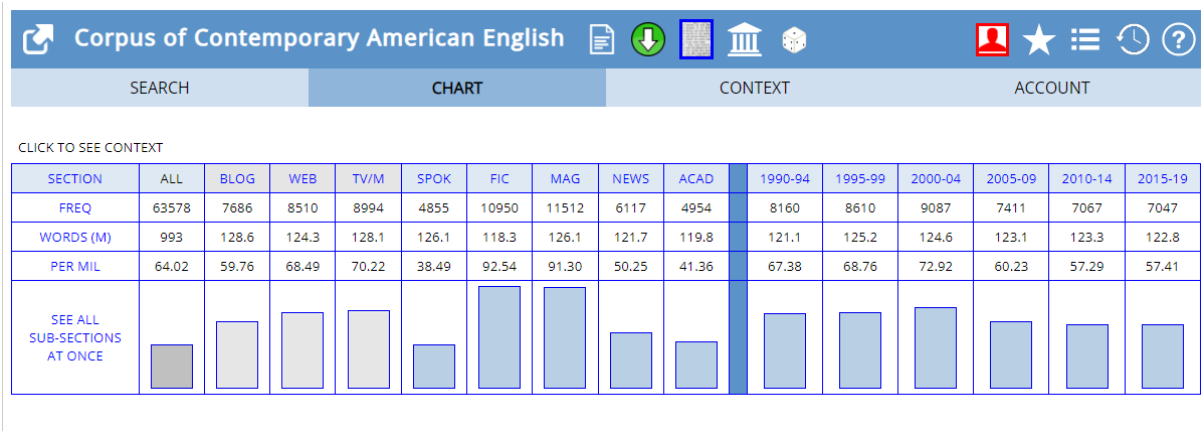
Word searched : Machine




Frequency:



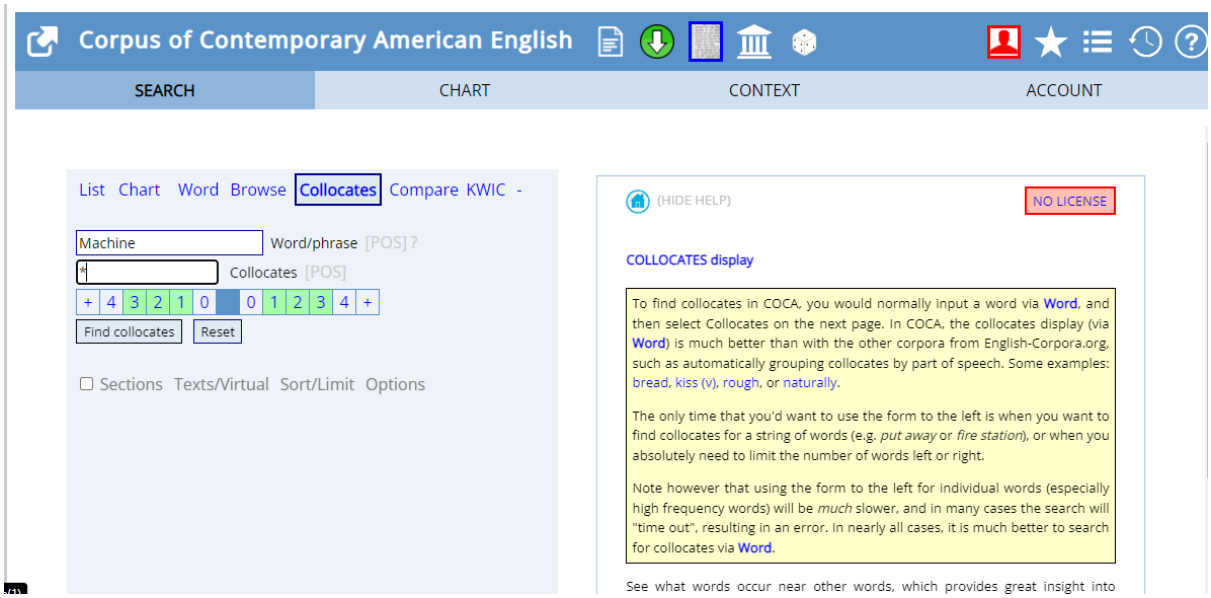

## 3.2 Chart – word – frequency; section; sub-section;




| SECTION | ALL | BLOG | WEB | TV/M | SPOK | FIC | MAG | NEWS | ACAD | 1990-94 | 1995-99 | 2000-04 | 2005-09 | 2010-14 | 2015-19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| FREQ | 63578 | 7686 | 8510 | 8994 | 4855 | 10950 | 11512 | 6117 | 4954 | 8160 | 8610 | 9087 | 7411 | 7067 | 7047 |
| WORDS (M) | 993 | 128.6 | 124.3 | 128.1 | 126.1 | 118.3 | 126.1 | 121.7 | 119.8 | 121.1 | 125.2 | 124.6 | 123.1 | 123.3 | 122.8 |
| PER MIL | 64.02 | 59.76 | 68.49 | 70.22 | 38.49 | 92.54 | 91.30 | 50.25 | 41.36 | 67.38 | 68.76 | 72.92 | 60.23 | 57.29 | 57.41 |

## 3.3 Collocate – Display three collocations on the left and right each of the word

Set to 3 collocations



Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | ACCOUNT

ON CLICK: CONTEXT TRANSLATE (??) ENTIRE PAGE GOOGLE IMAGE PRON/VIDEO BOOK (HELP)

| HELP | | | | FREQ |
|---|---|---|---|---|
| 1 | | | THE | 29786 |
| 2 | | | . | 20784 |
| 3 | | | A | 19643 |
| 4 | | | , | 17478 |
| 5 | | | AND | 9225 |
| 6 | | | TO | 6965 |
| 7 | | | OF | 5838 |
| 8 | | | THAT | 5391 |
| 9 | | | IN | 5003 |
| 10 | | | " | 4847 |
| 11 | | | ON | 4675 |
| 12 | | | 'S | 4020 |
| 13 | | | IS | 3838 |
| 14 | | | WITH | 3287 |
| 15 | | | IT | 2732 |
| 16 | | | FOR | 2544 |
| 17 | | | I | 2530 |
| 18 | | | THIS | 2527 |
| 19 | | | AN | 2296 |
| 20 | | | TIME | 2268 |
| 21 | | | WAS | 2105 |
| 22 | | | ? | 2076 |
| 23 | | | GUN | 2041 |
| 24 | | | YOU | 2017 |
| 25 | | | ANSWERING | 1894 |

Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | ACCOUNT

| | | | | |
|---|---|---|---|---|
| 35 | | | INTO | 1450 |
| 36 | | | WASHING | 1408 |
| 37 | | | HE | 1400 |
| 38 | | | ( | 1307 |
| 39 | | | : | 1306 |
| 40 | | | BUT | 1243 |
| 41 | | | LIKE | 1137 |
| 42 | | | VENDING | 1087 |
| 43 | | | HAVE | 1048 |
| 44 | | | CAN | 1034 |
| 45 | | | - | 1012 |
| 46 | | | UP | 1009 |
| 47 | | | ) | 993 |
| 48 | | | NOT | 981 |
| 49 | | | HAS | 936 |
| 50 | | | LEARNING | 926 |
| 51 | | | BE | 924 |
| 52 | | | SEWING | 848 |
| 53 | | | HAD | 843 |
| 54 | | | WHICH | 836 |
| 55 | | | HER | 834 |
| 56 | | | OUT | 831 |
| 57 | | | THEY | 830 |
| 58 | | | ONE | 814 |
| 59 | | | WE | 790 |
| 60 | | | FAX | 772 |
| 61 | | | WILL | 755 |
| 62 | | | ! | 743 |

Corpus of Contemporary American English

SEARCH | FREQUENCY | CONTEXT | ACCOUNT

| | | | | |
|---|---|---|---|---|
| 76 | | | NO | 602 |
| 77 | | | COULD | 592 |
| 78 | | | SHE | 585 |
| 79 | | | ' | 574 |
| 80 | | | ITS | 571 |
| 81 | | | USE | 571 |
| 82 | | | DO | 562 |
| 83 | | | OFF | 506 |
| 84 | | | ALL | 505 |
| 85 | | | ; | 496 |
| 86 | | | THEN | 491 |
| 87 | | | THROUGH | 489 |
| 88 | | | GET | 465 |
| 89 | | | USING | 445 |
| 90 | | | ABOUT | 422 |
| 91 | | | BIG | 422 |
| 92 | | | SOME | 420 |
| 93 | | | OUR | 416 |
| 94 | | | USED | 408 |
| 95 | | | MAN | 405 |
| 96 | | | AGAINST | 404 |
| 97 | | | GOT | 403 |
| 98 | | | THAN | 403 |
| 99 | | | POLITICAL | 390 |
| 100 | | | MORE | 389 |
| | | | TOTAL | 228815 |

SEE MORE THAN 100 WORDS/PHRASES

0.930 seconds