

recruit

June 7, 2024

DP

```
[59]: !pip install ja-ginza
```

```
Requirement already satisfied: ja-ginza in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (5.2.0)
Requirement already satisfied: spacy<4.0.0,>=3.4.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from ja-ginza)
(3.7.5)
Requirement already satisfied: sudachipy<0.7.0,>=0.6.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from ja-ginza)
(0.6.8)
Requirement already satisfied: sudachidict-core>=20210802 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from ja-ginza)
(20240409)
Requirement already satisfied: ginza<5.3.0,>=5.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from ja-ginza)
(5.2.0)
Requirement already satisfied: plac>=1.3.3 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
ginza<5.3.0,>=5.2.0->ja-ginza) (1.4.3)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (8.2.4)
```

Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (1.1.3)

Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.4.8)

Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.0.10)

Requirement already satisfied: weasel<0.5.0,>=0.1.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (0.4.1)

Requirement already satisfied: typer<1.0.0,>=0.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (0.12.3)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (4.66.2)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.31.0)

Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (2.7.0)

Requirement already satisfied: jinja2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (3.1.3)

Requirement already satisfied: setuptools in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (69.5.1)

Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (23.2)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (3.4.0)

Requirement already satisfied: numpy>=1.19.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
spacy<4.0.0,>=3.4.4->ja-ginza) (1.26.4)

Requirement already satisfied: language-data>=1.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
langcodes<4.0.0,>=3.2.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.2.0)

Requirement already satisfied: annotated-types>=0.4.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (0.6.0)

Requirement already satisfied: pydantic-core==2.18.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (2.18.1)

Requirement already satisfied: typing-extensions>=4.6.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy<4.0.0,>=3.4.4->ja-ginza) (4.11.0)

Requirement already satisfied: charset-normalizer<4,>=2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.26.18)

Requirement already satisfied: certifi>=2017.4.17 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy<4.0.0,>=3.4.4->ja-ginza) (2024.2.2)

Requirement already satisfied: blis<0.8.0,>=0.7.8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy<4.0.0,>=3.4.4->ja-ginza) (0.7.11)

Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy<4.0.0,>=3.4.4->ja-ginza) (0.1.5)

Requirement already satisfied: click>=8.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (8.1.7)

Requirement already satisfied: shellingham>=1.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.5.4)

Requirement already satisfied: rich>=10.11.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (13.7.1)

Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza) (0.18.1)

Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza) (7.0.4)

Requirement already satisfied: MarkupSafe>=2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
jinja2->spacy<4.0.0,>=3.4.4->ja-ginza) (2.1.5)

Requirement already satisfied: marisa-trie>=0.7.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from language-
data>=1.2->langcodes<4.0.0,>=3.2.0->spacy<4.0.0,>=3.4.4->ja-ginza) (1.2.0)

Requirement already satisfied: markdown-it-py>=2.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (3.0.0)

Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza) (2.17.2)

```
Requirement already satisfied: wrapt in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy<4.0.0,>=3.4.4->ja-ginza)
(1.16.0)
Requirement already satisfied: mdurl~=0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from markdown-it-
py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy<4.0.0,>=3.4.4->ja-ginza)
(0.1.2)
```

0.1 1.

```
[60]: !pip install spacy
      !pip install sklearn
      !pip install python-dp
```

```
Requirement already satisfied: spacy in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (3.7.5)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.0.9)
Requirement already satisfied: thinc<8.3.0,>=8.2.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(8.2.4)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in
```

/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(0.12.3)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(4.66.2)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.31.0)

Requirement already satisfied: pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(2.7.0)

Requirement already satisfied: jinja2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.1.3)

Requirement already satisfied: setuptools in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(69.5.1)

Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(23.2)

Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(3.4.0)

Requirement already satisfied: numpy>=1.19.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from spacy)
(1.26.4)

Requirement already satisfied: language-data>=1.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)

Requirement already satisfied: annotated-types>=0.4.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (0.6.0)

Requirement already satisfied: pydantic-core==2.18.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (2.18.1)

Requirement already satisfied: typing-extensions>=4.6.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pydantic!=1.8,!1.8.1,<3.0.0,>=1.7.4->spacy) (4.11.0)

Requirement already satisfied: charset-normalizer<4,>=2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (1.26.18)

Requirement already satisfied: certifi>=2017.4.17 in

```

/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
requests<3.0.0,>=2.13.0->spacy) (2024.2.2)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy) (0.7.11)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
thinc<8.3.0,>=8.2.2->spacy) (0.1.5)
Requirement already satisfied: click>=8.0.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (8.1.7)
Requirement already satisfied: shellingham>=1.3.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (1.5.4)
Requirement already satisfied: rich>=10.11.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
typer<1.0.0,>=0.3.0->spacy) (13.7.1)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (0.18.1)
Requirement already satisfied: smart-open<8.0.0,>=5.2.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
weasel<0.5.0,>=0.1.0->spacy) (7.0.4)
Requirement already satisfied: MarkupSafe>=2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
jinja2->spacy) (2.1.5)
Requirement already satisfied: marisa-trie>=0.7.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from language-
data>=1.2->langcodes<4.0.0,>=3.2.0->spacy) (1.2.0)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (2.17.2)
Requirement already satisfied: wrapt in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from smart-
open<8.0.0,>=5.2.1->weasel<0.5.0,>=0.1.0->spacy) (1.16.0)
Requirement already satisfied: mdurl~=0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from markdown-it-
py>=2.2.0->rich>=10.11.0->typer<1.0.0,>=0.3.0->spacy) (0.1.2)
Collecting sklearn
  Using cached sklearn-0.0.post12.tar.gz (2.6 kB)
  Installing build dependencies ... done
  Getting requirements to build wheel ... error
  error: subprocess-exited-with-error

  × Getting requirements to build wheel did not run

```

successfully.

exit code: 1

> [15 lines of output]

The 'sklearn' PyPI package is deprecated, use 'scikit-learn' rather than 'sklearn' for pip commands.

Here is how to fix this error in the main use cases:

- use 'pip install scikit-learn' rather than 'pip install sklearn'
- replace 'sklearn' by 'scikit-learn' in your pip requirements files (requirements.txt, setup.py, setup.cfg, Pipfile, etc ...)
- if the 'sklearn' package is used by one of your dependencies, it would be great if you take some time to track which package uses 'sklearn' instead of 'scikit-learn' and report it to their issue tracker
- as a last resort, set the environment variable SKLEARN_ALLOW_DEPRECATED_SKLEARN_PACKAGE_INSTALL=True to avoid this error

More information is available at

<https://github.com/scikit-learn/sklearn-pypi-package>

[end of output]

note: This error originates from a subprocess, and is likely not a problem with pip.

error: subprocess-exited-with-error

× Getting requirements to build wheel did not run successfully.

exit code: 1

> See above for output.

note: This error originates from a subprocess, and is likely not a problem with pip.

Requirement already satisfied: python-dp in /home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (1.1.4)

0.2 2.

```
[61]: import pandas as pd

data_path = './data/reviews_with_sentiment.csv'

df = pd.read_csv(data_path)
df
```

```
[61]:
```

	review	sentiment
0		neutral
1		neutral
2		positive
3		positive
4		positive
...
5548		negative
5549		negative
5550		neutral
5551		negative
5552	...	neutral

[5553 rows x 2 columns]

```
[62]: import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
import spacy
import itertools
from typing import List, Tuple
from pydp.algorithms.laplacian import Count
import math

# spaCy
nlp = spacy.load('ja_ginza')

#
data_path = './data/reviews_with_sentiment.csv'
df = pd.read_csv(data_path)

#
POS = ['ADJ', 'ADV', 'INTJ', 'PROPN', 'NOUN', 'VERB']
MAX_TERMS_IN_DOC = 5
NGRAM = 1
MAX_DF = 1.0
MIN_DF = 0.0
NUM_VOCAB = 10000
TOP_K = 20

def flatten(*lists) -> list:
    res = []
    for l in list(itertools.chain.from_iterable(lists)):
        for e in l:
            res.append(e)
```



```

    return res

def remove_duplicates(l: List[Tuple[str, float]]) -> List[Tuple[str, float]]:
    d = {}
    for e in l:
        d[e[0]] = e[1]
    return list(d.items())

def preprocess_for_private_counts(tf: np.ndarray) -> List[np.ndarray]:
    repeated_words = []
    for i, term in enumerate(tf):
        repeated_words.append(np.repeat(i, term))
    return repeated_words

def cal_private_count(epsilon: float, max_partition_contributed: float,
    ↪max_contributions_per_partition: float, repeated_words: List[np.ndarray]) ->
    ↪List[int]:
    private_counts = []
    for repeated_word in repeated_words:
        counter = Count(epsilon, max_partition_contributed,
    ↪max_contributions_per_partition)
        count = counter.quick_result(repeated_word)
        private_counts.append(count)
    return private_counts

def top_k_words_and_counts(k: int, tf: np.ndarray, vocab: dict) ->
    ↪List[Tuple[str, int]]:
    indices_topk = np.argsort(tf)[:,-1][:k]
    reverse_vocab = {vocab[key]: key for key in vocab.keys()}
    words = [reverse_vocab[i] for i in indices_topk]
    counts = [tf[i] for i in indices_topk]
    return list(zip(words, counts))

#
df["doc"] = [nlp(review) for review in df["review"]]

```

0.3 3. Bag-of-Words

```

[63]: # Bag-of-Words
bows = {}
cvs = {}

for sentiment in df["sentiment"].unique():
    tokens = []
    for doc in df[df["sentiment"] == sentiment]["doc"]:
        similarities = [(token.similarity(doc), token.lemma_) for token in doc
    ↪if token.pos_ in POS]

```

```

        similarities = remove_duplicates(similarities)
        similarities = sorted(similarities, key=lambda sim: sim[1],
reverse=True)[:MAX_TERMS_IN_DOC]
        tokens.append([similarity[1] for similarity in similarities])

    cv = CountVectorizer(ngram_range=(1, NGRAM), max_df=MAX_DF, min_df=MIN_DF,
max_features=NUM_VOCAB)
    bows[sentiment] = cv.fit_transform(flatten(tokens)).toarray()
    cvs[sentiment] = cv

```

/tmp/ipykernel_931337/2647833454.py:8: UserWarning: [W008] Evaluating Token.similarity based on empty vectors.

```

    similarities = [(token.similarity(doc), token.lemma_) for token in doc if
token.pos_ in POS]

```

```

[64]: !pip install numpy
      !pip install matplotlib
      !pip install seaborn

```

```

Requirement already satisfied: numpy in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (1.26.4)
Requirement already satisfied: matplotlib in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (3.8.4)
Requirement already satisfied: contourpy>=1.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.2.1)
Requirement already satisfied: cycler>=0.10 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(4.51.0)
Requirement already satisfied: kiwisolver>=1.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.4.5)
Requirement already satisfied: numpy>=1.21 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(1.26.4)
Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(23.2)
Requirement already satisfied: pillow>=8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(10.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib)
(3.1.2)

```

Requirement already satisfied: python-dateutil>=2.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from matplotlib
(2.9.0.post0))

Requirement already satisfied: six>=1.5 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)

Requirement already satisfied: seaborn in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (0.13.2)

Requirement already satisfied: numpy!=1.24.0,>=1.20 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(1.26.4)

Requirement already satisfied: pandas>=1.2 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(2.2.2)

Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from seaborn)
(3.8.4)

Requirement already satisfied: contourpy>=1.0.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.2.1)

Requirement already satisfied: cycler>=0.10 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)

Requirement already satisfied: kiwisolver>=1.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)

Requirement already satisfied: packaging>=20.0 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (23.2)

Requirement already satisfied: pillow>=8 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)

Requirement already satisfied: pyparsing>=2.3.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (3.1.2)

Requirement already satisfied: python-dateutil>=2.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pandas>=1.2->seaborn) (2024.1)

Requirement already satisfied: tzdata>=2022.7 in
/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from
pandas>=1.2->seaborn) (2024.1)

Requirement already satisfied: six>=1.5 in

/home/jun/.pyenv/versions/3.11.8/lib/python3.11/site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)

0.4 4.

```
[65]: #
vocabs = {}
term_frequencies = {}

for sentiment in df["sentiment"].unique():
    bow = bows[sentiment]
    cv = cvs[sentiment]

    vocab = cv.vocabulary_
    term_frequency = np.sum(bow, axis=0)
    vocabs[sentiment] = vocab
    term_frequencies[sentiment] = term_frequency

    indices_topk = np.argsort(term_frequency)[::-1][:TOP_K]
    bow_topk = np.take(bow, indices_topk, axis=1)
    reverse_vocab = {vocab[k]: k for k in vocab.keys()}
    words = [reverse_vocab[i] for i in indices_topk]

    print(sentiment, ":")
    for w, c in zip(words, term_frequency[indices_topk]):
        print(w, ":", c)
```

neutral :

```
: 216
: 93
: 67
: 60
: 59
: 50
: 47
: 45
: 43
: 43
: 34
: 32
: 30
: 27
: 27
: 25
: 25
: 23
: 22
: 21
```

positive :

: 463
: 315
: 206
: 198
: 193
: 181
: 175
: 144
: 141
: 140
: 124
: 106
: 102
: 99
: 96
: 90
: 90
: 87
: 84
: 80

negative :

: 99
: 97
: 48
: 42
: 39
: 36
: 34
: 33
: 27
: 24
: 22
: 19
: 16
: 16
: 16
: 16
: 15
: 15
: 15
: 14

0.5 5.

```
[66]: import numpy as np
import matplotlib.pyplot as plt

#
def clamp_values(values: np.ndarray, LAMBDA: float) -> np.ndarray:
    return np.clip(values, -LAMBDA, LAMBDA)

#
def plot_histogram(values: np.ndarray, title: str):
    plt.hist(values, bins=30, alpha=0.7, edgecolor='black')
    plt.title(title)
    plt.xlabel('Value')
    plt.ylabel('Frequency')
    plt.show()

#
epsilons = [0.01, 0.05, 0.1, 0.3, 0.7, 1.0, 2.0, 3.0, 7.0, 10.0]
MAX_DUPLICATED_TERMS = 1
LAMBDA = 500 #

#
results = {sentiment: {"no DP": None, **{eps: None for eps in epsilons}} for
    sentiment in df["sentiment"].unique()}

for eps in epsilons:
    print(" : ", eps)
    for sentiment in df["sentiment"].unique():
        repeated_words =
    preprocess_for_private_counts(term_frequencies[sentiment])
        private_counts = cal_private_count(eps, MAX_TERMS_IN_DOC,
    MAX_DUPLICATED_TERMS, repeated_words)

        #
        plot_histogram(private_counts, f'Original Values (={eps},
    sentiment={sentiment})')

        #
        -LAMBDA LAMBDA
        clamped_counts = clamp_values(private_counts, LAMBDA)

        #
        plot_histogram(clamped_counts, f'Clamped Values (={eps},
    sentiment={sentiment})')

        #
```

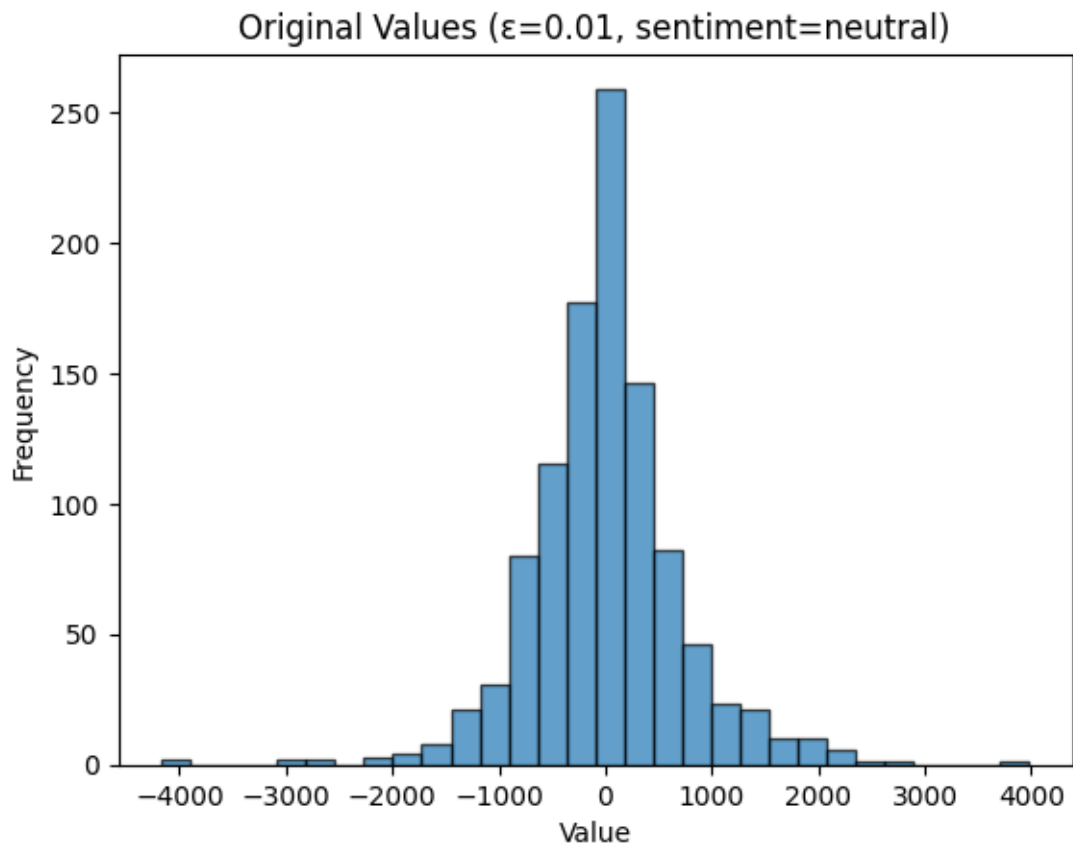
```

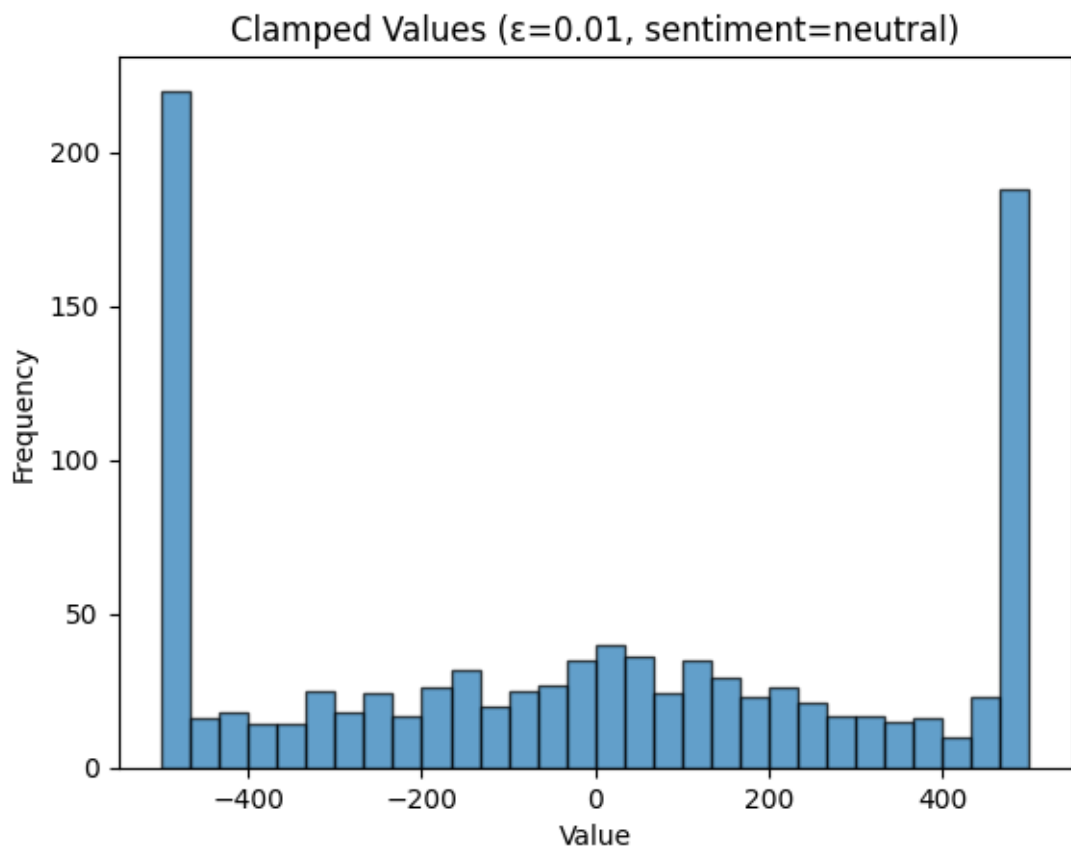
    bias_corrected_counts = [count - LAMBDA * np.exp(-eps) for count in
↪clamped_counts]

    words_and_counts = top_k_words_and_counts(TOP_K, bias_corrected_counts,
↪vocabs[sentiment])
    results[sentiment][eps] = words_and_counts
    print(sentiment, ":")
    print(words_and_counts)

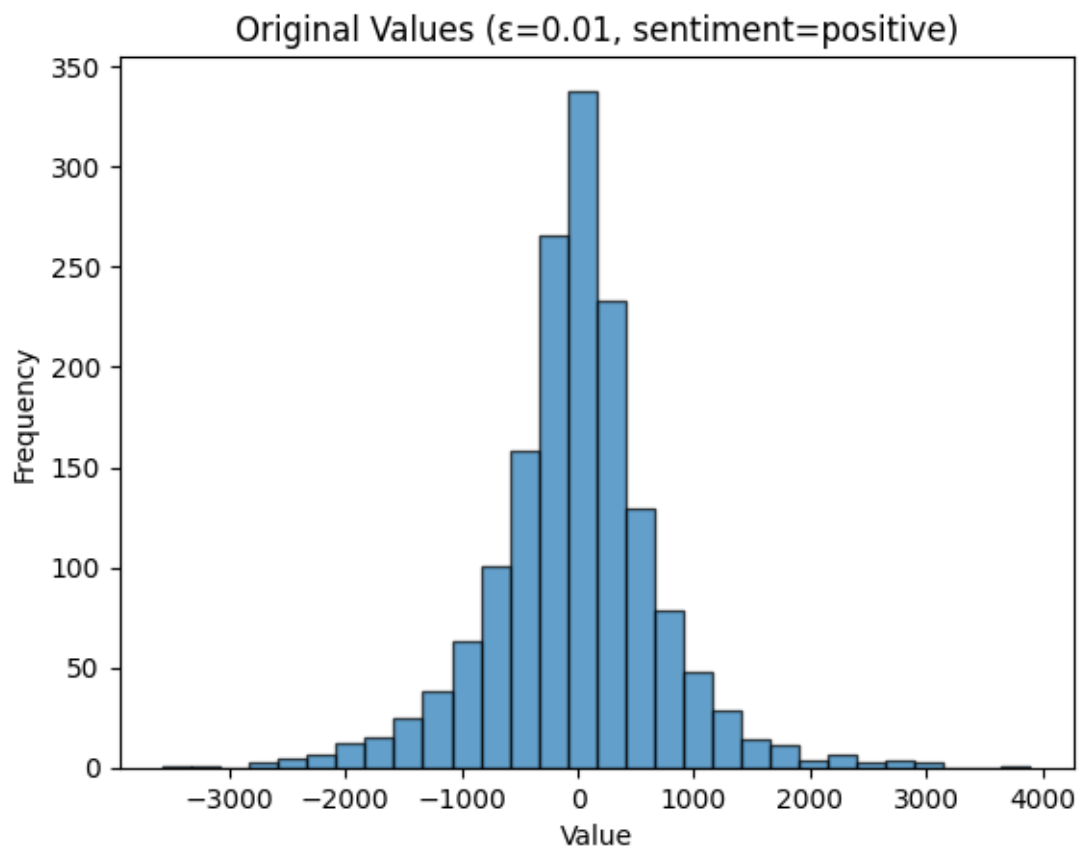
```

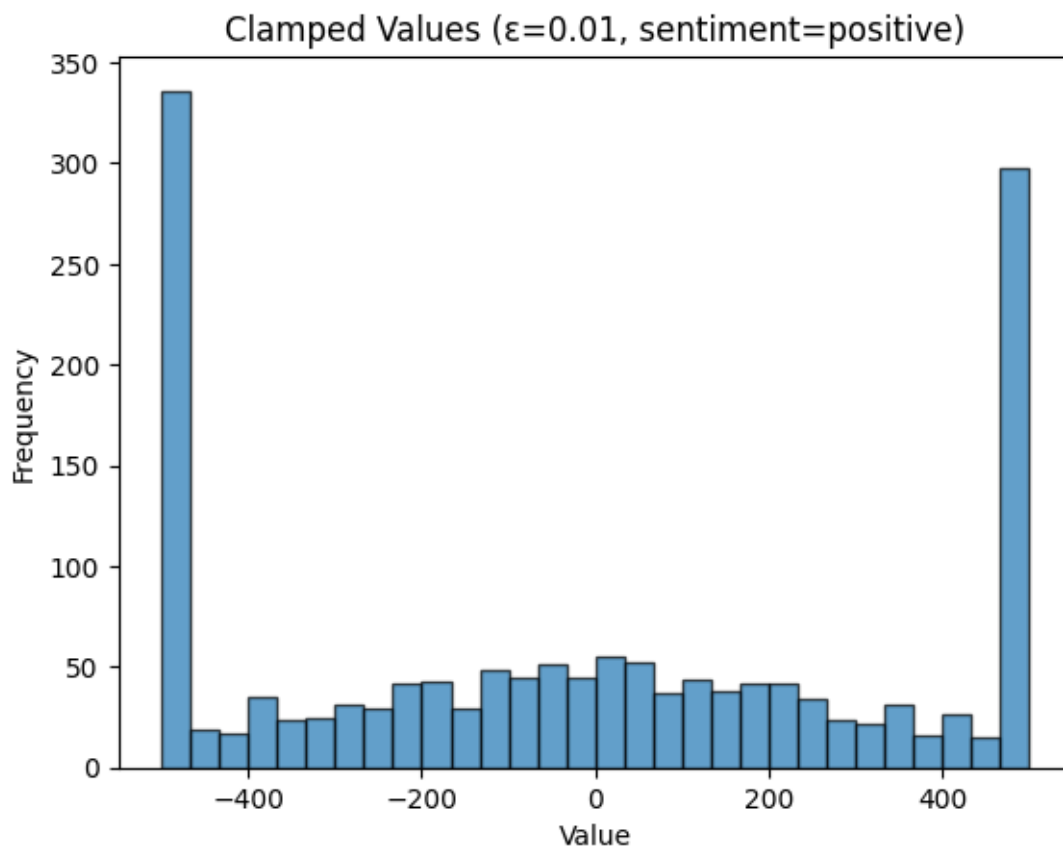
: 0.01





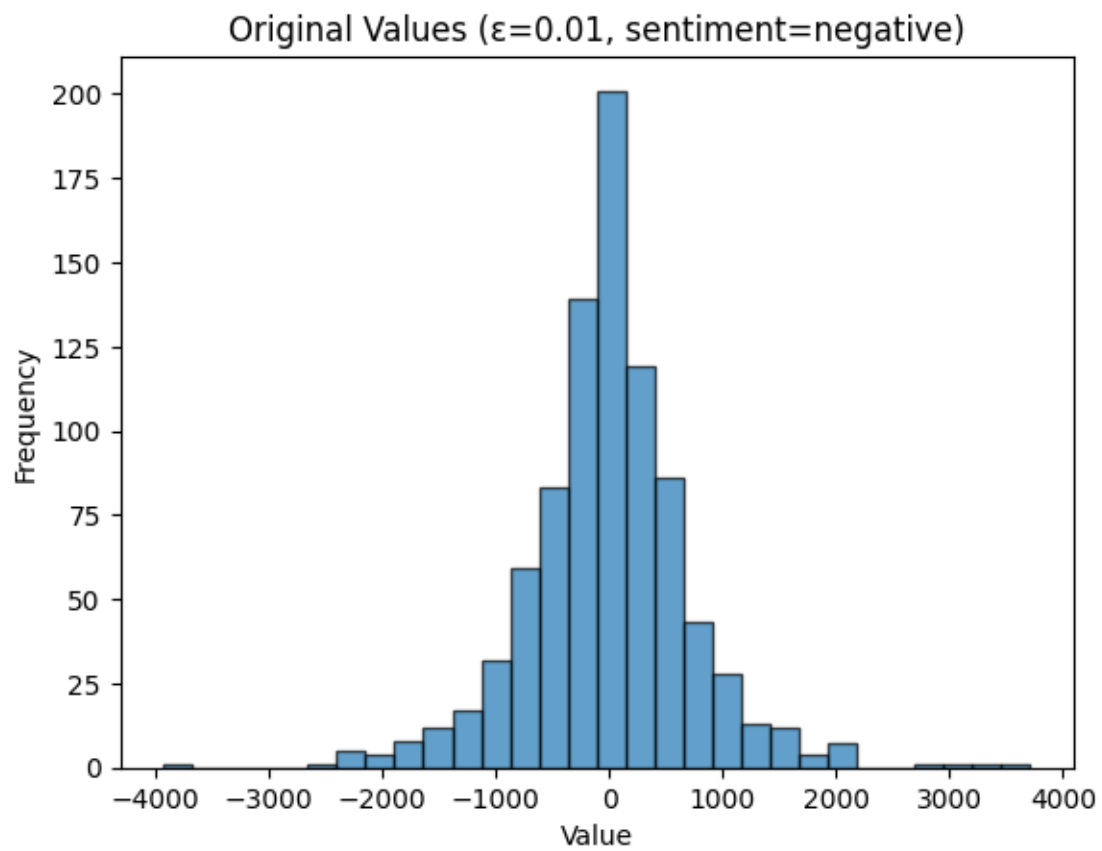
```
neutral :
[(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949)]
```

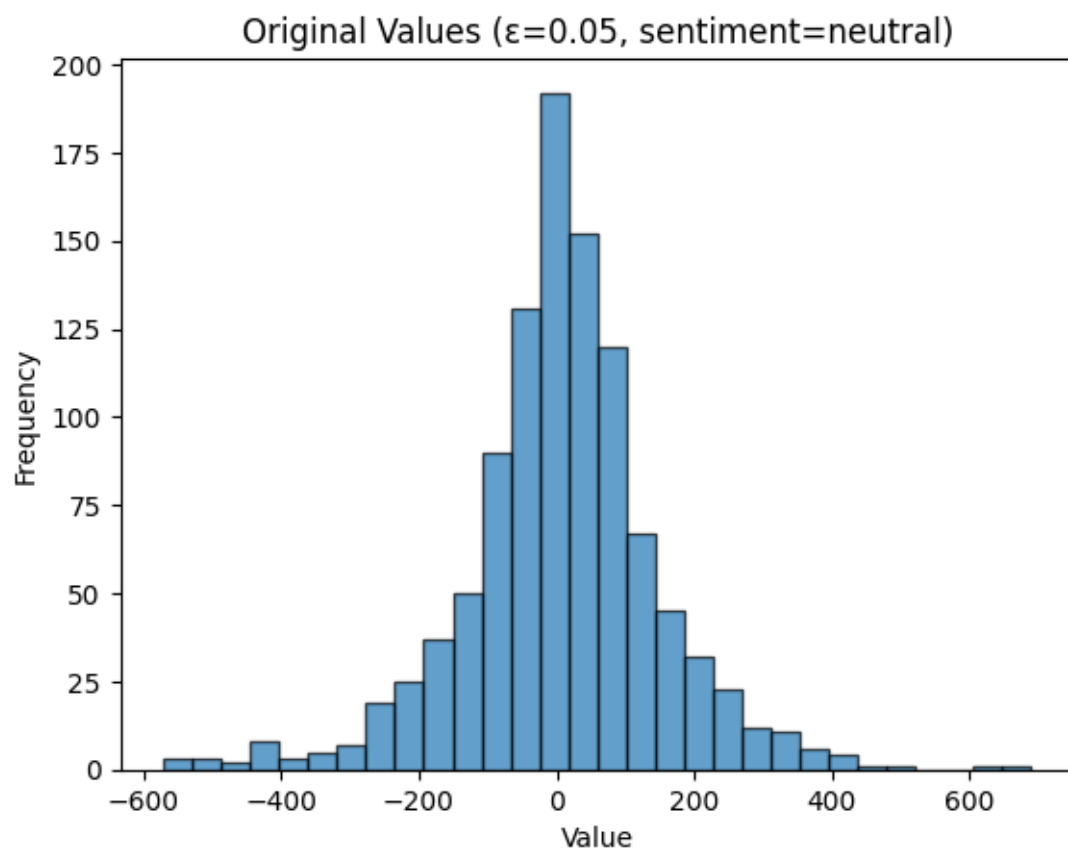


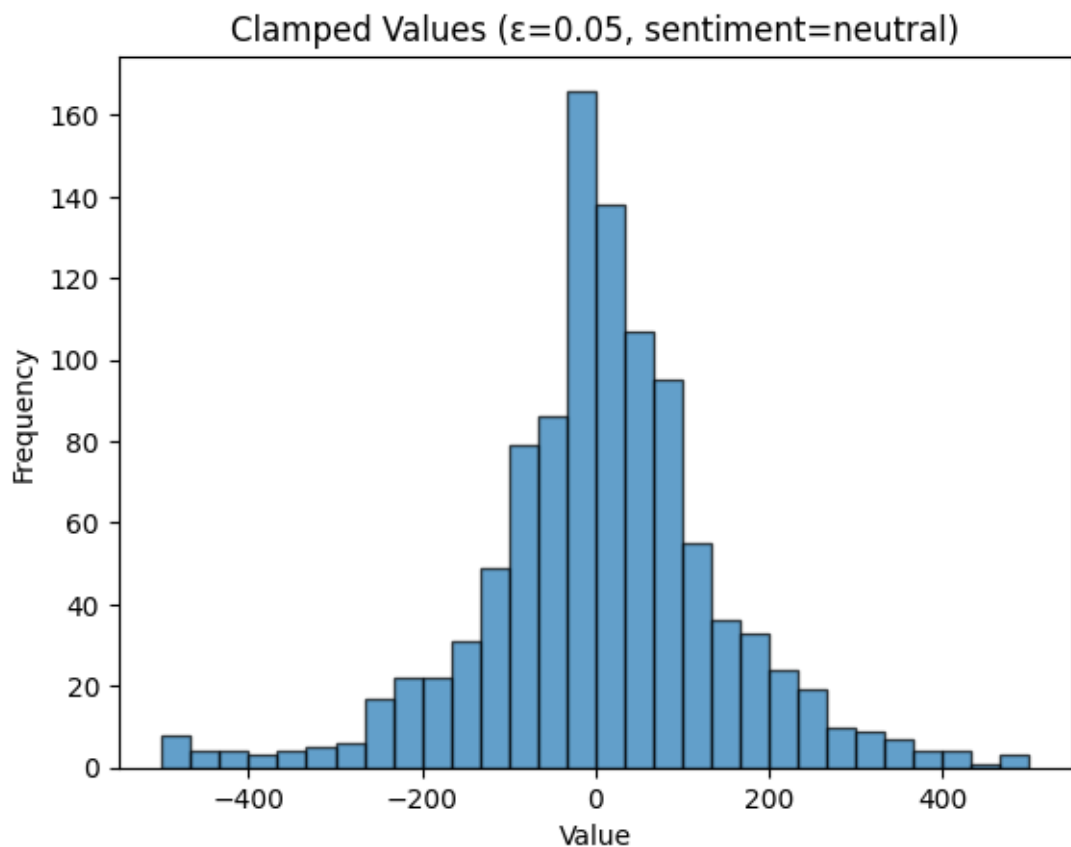
positive :

```
[(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949), (' ', 4.975083125415949),
(' ', 4.975083125415949), (' ', 4.975083125415949), (' ',
4.975083125415949), (' ', 4.975083125415949)]
```



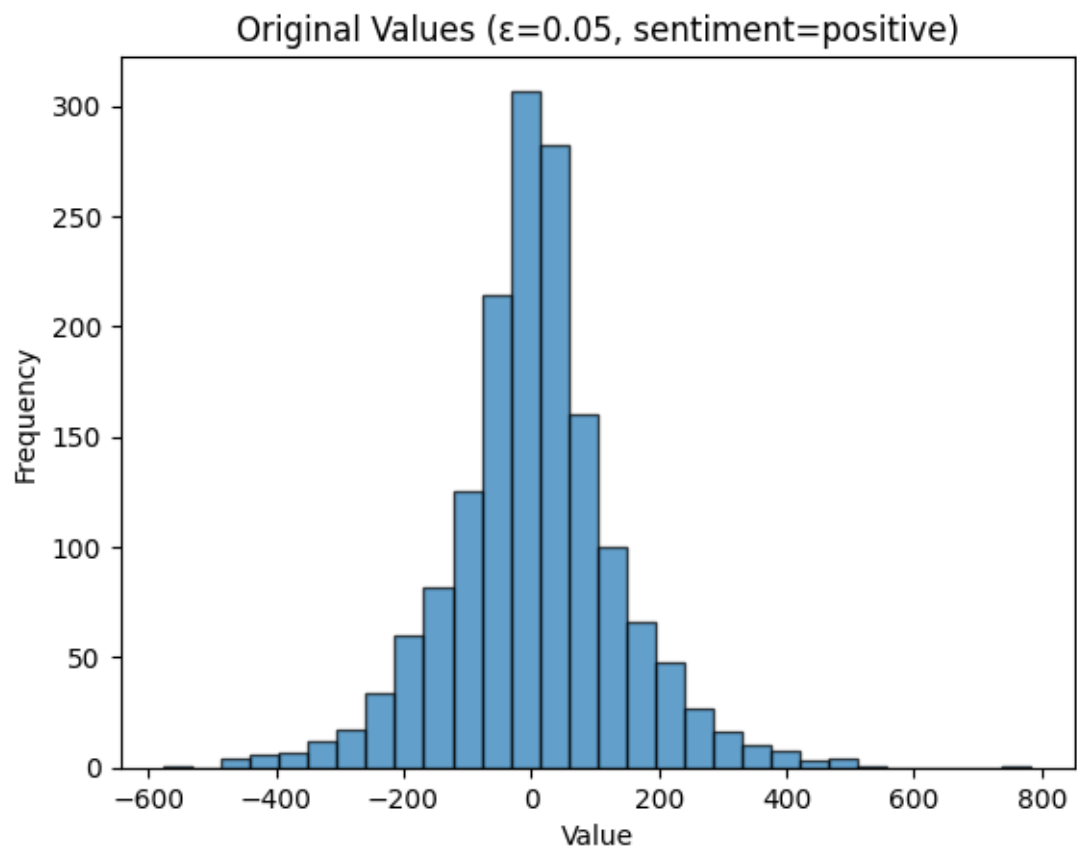
20

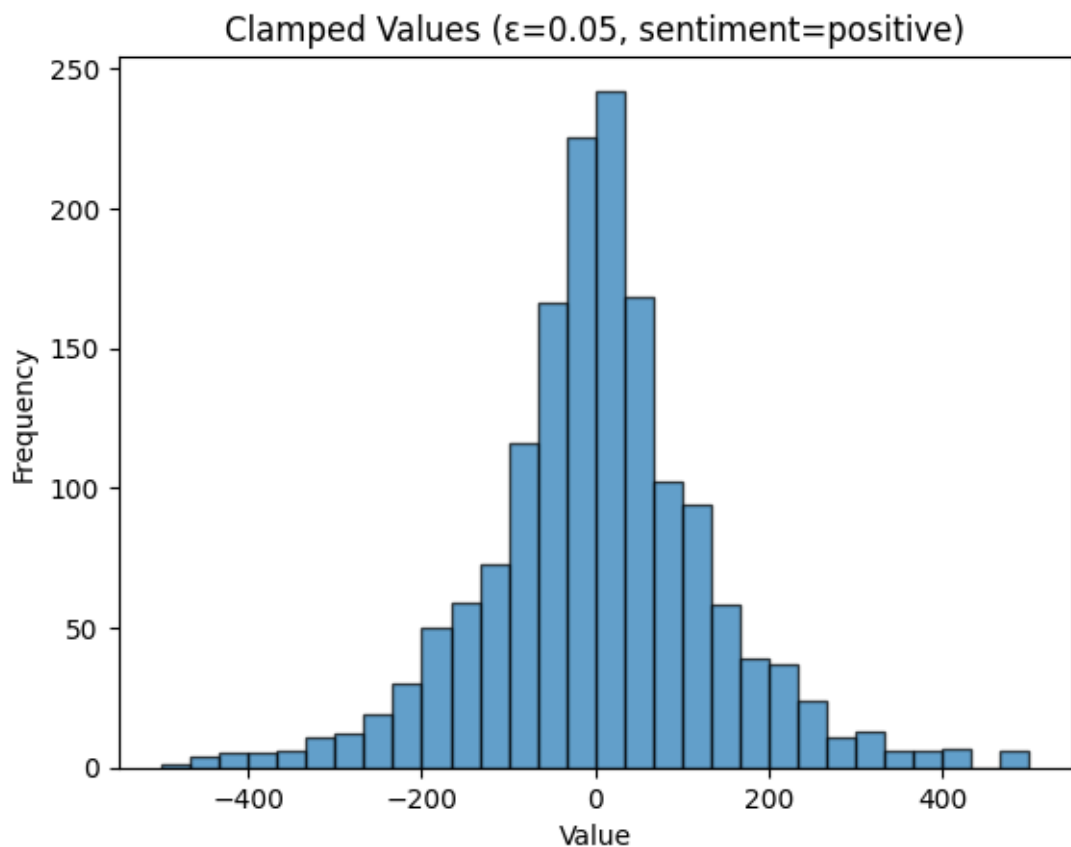




neutral :

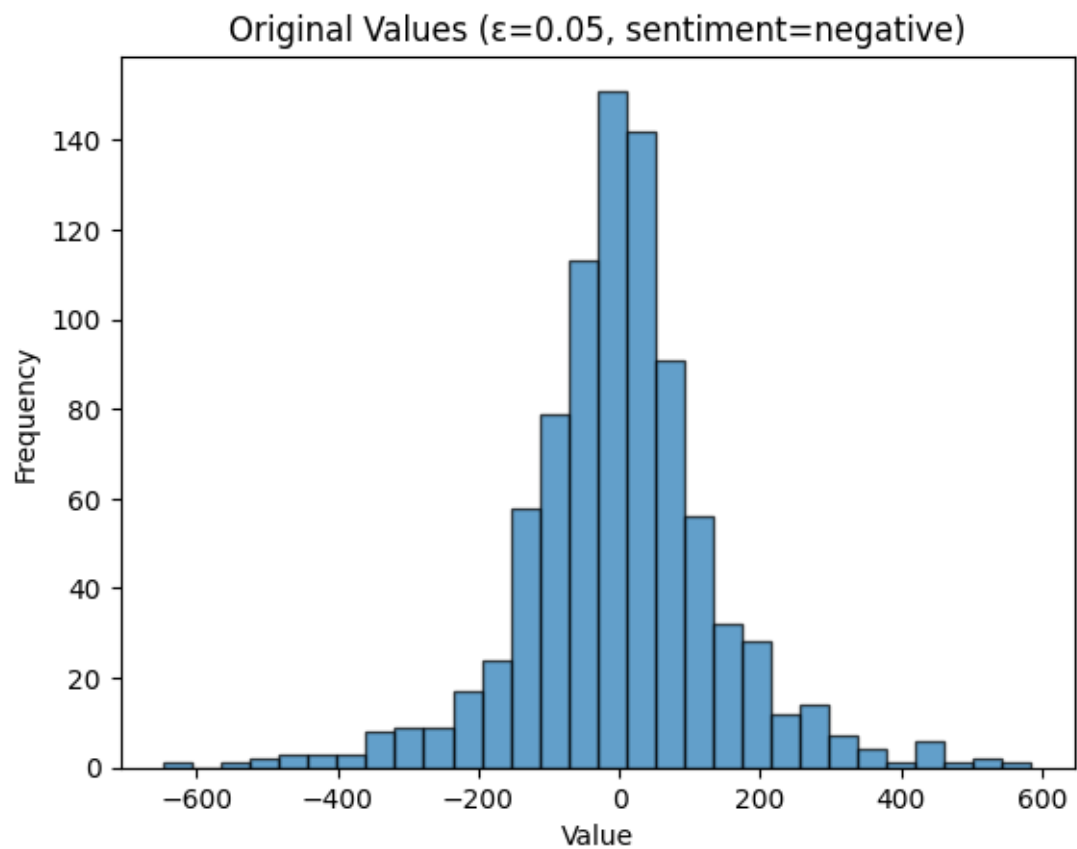
```
[(' ', 24.385287749643), (' ', 24.385287749643), (' ', 24.385287749643),
(' ', -32.614712250357), (' ', -42.614712250357), (' ', -57.614712250357),
(' ', -59.614712250357), (' ', -72.614712250357), (' ', -91.614712250357),
(' ', -92.614712250357), (' ', -99.614712250357), (' ', -108.614712250357),
(' ', -119.614712250357), (' ', -119.614712250357), (' ',
-123.614712250357), (' ', -126.614712250357), (' ', -135.614712250357), (' ',
-135.614712250357), (' ', -140.614712250357), (' ', -149.614712250357)]
```

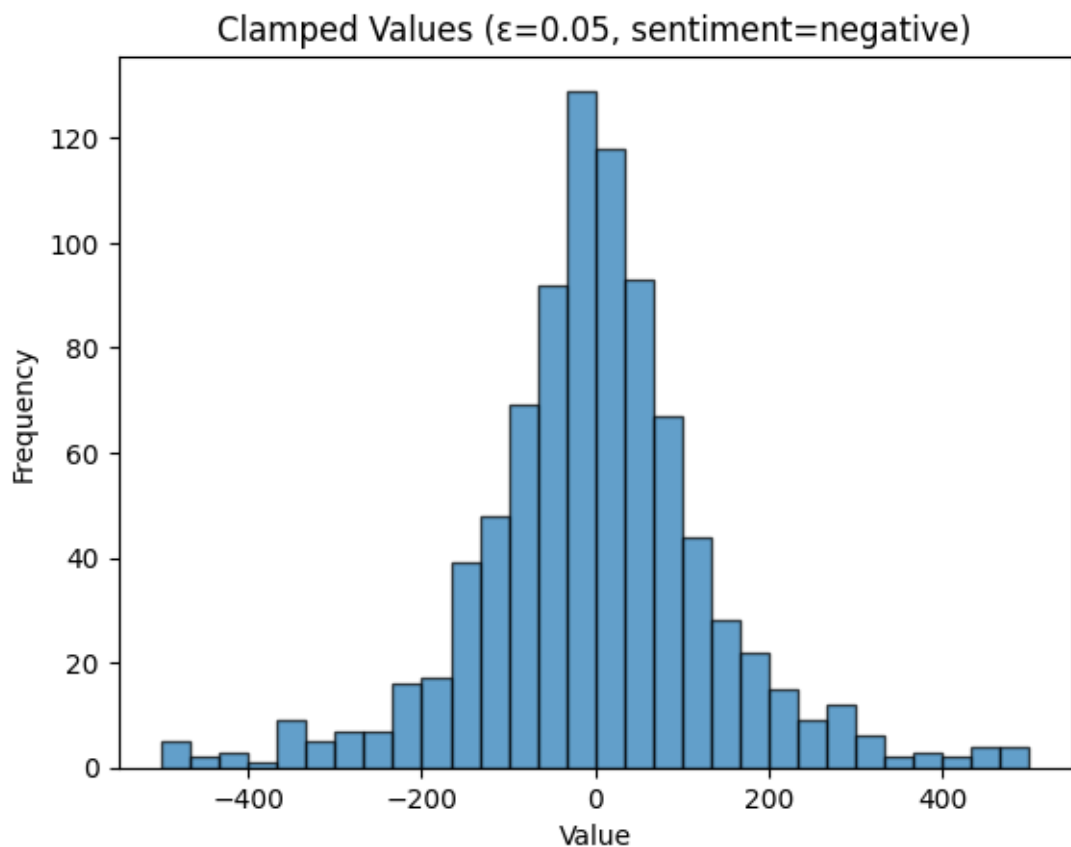




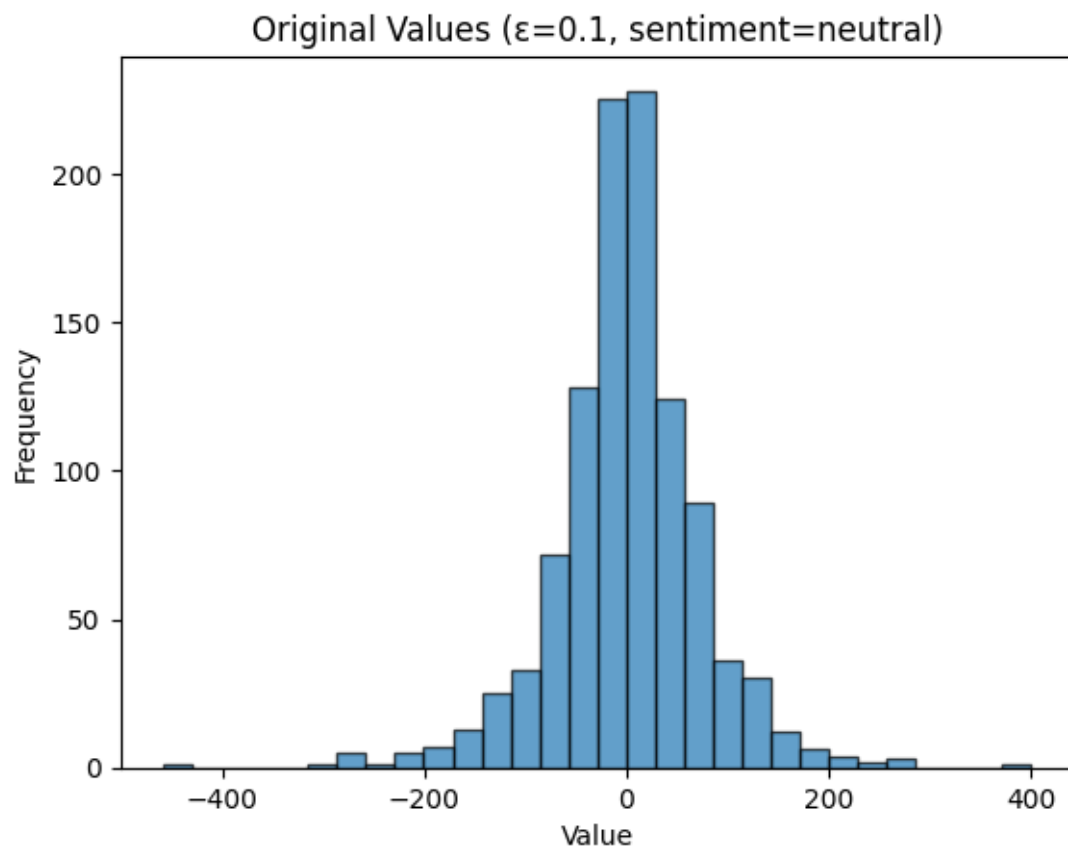
positive :

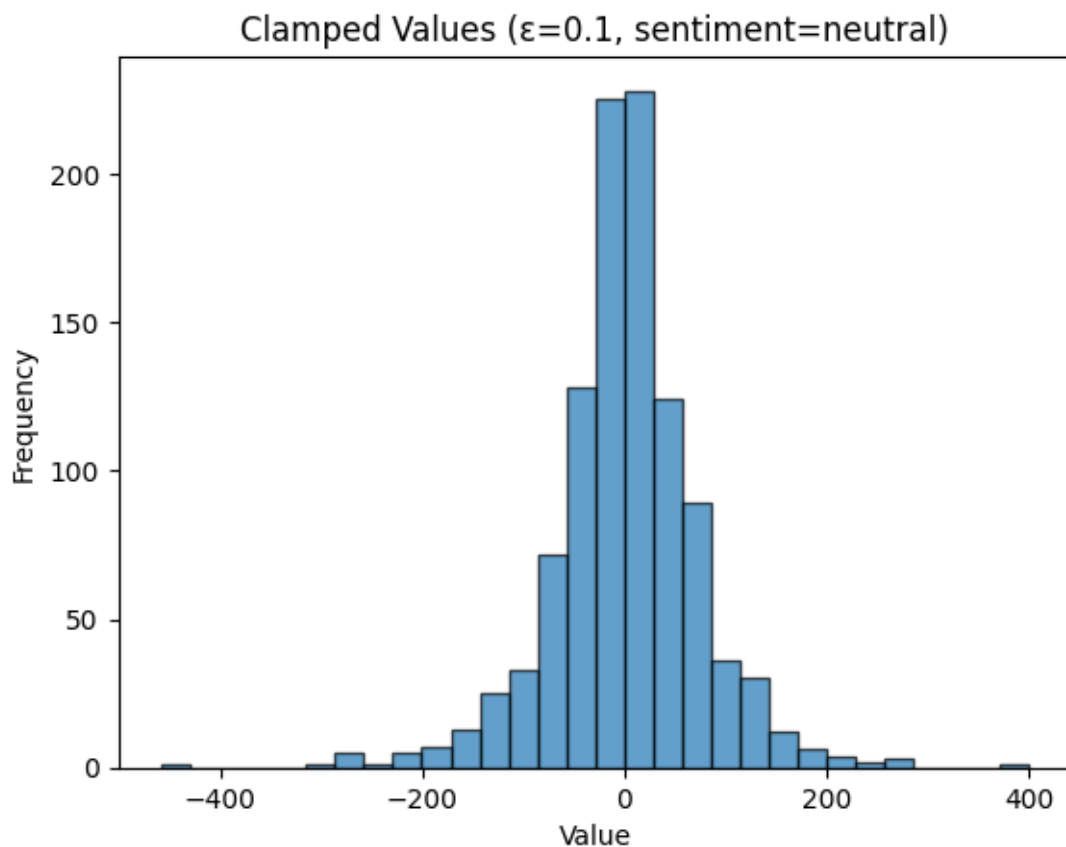
```
[(' ', 24.385287749643), (' ', 24.385287749643), (' ', 24.385287749643),
(' ', 14.385287749642998), (' ', 14.385287749642998), (' ',
-3.6147122503570017), (' ', -42.614712250357), (' ', -42.614712250357),
(' ', -49.614712250357), (' ', -59.614712250357), (' ', -66.614712250357),
(' ', -67.614712250357), (' ', -73.614712250357), (' ',
-78.614712250357), (' ', -84.614712250357), (' ', -89.614712250357), (' ',
-98.614712250357), (' ', -100.614712250357), (' ', -100.614712250357), (' ',
-109.614712250357)]
```

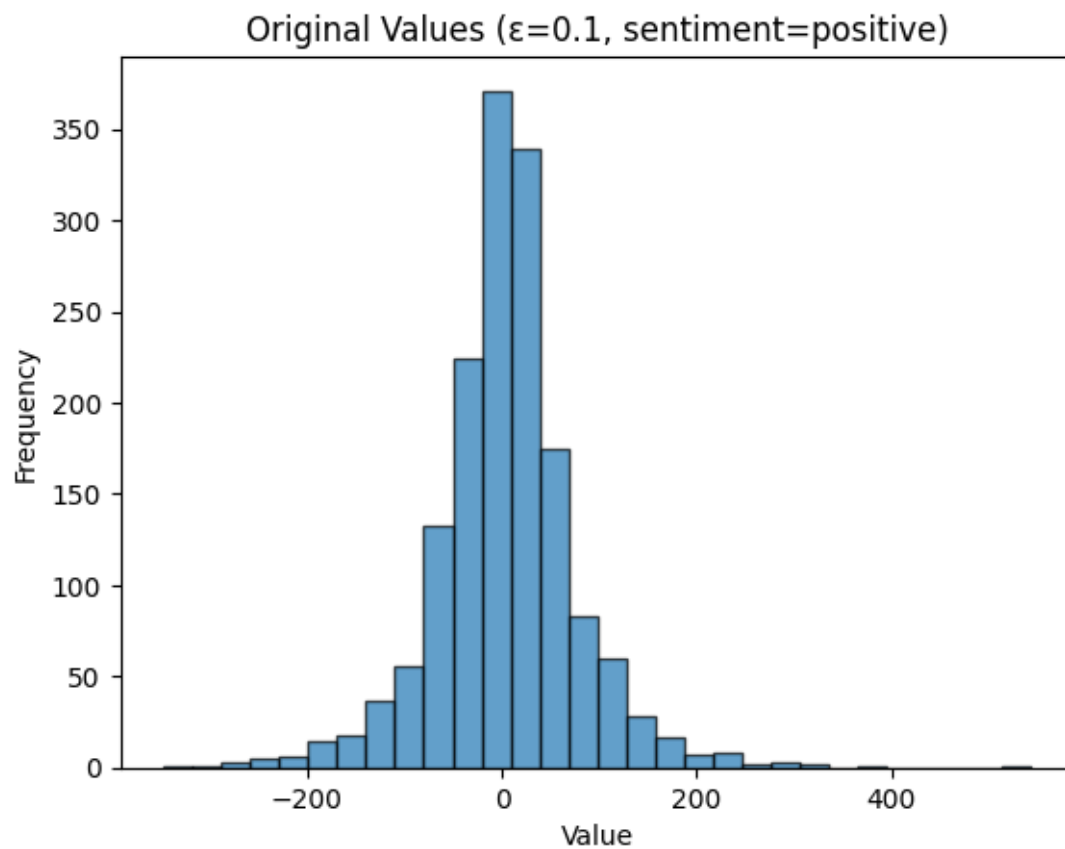


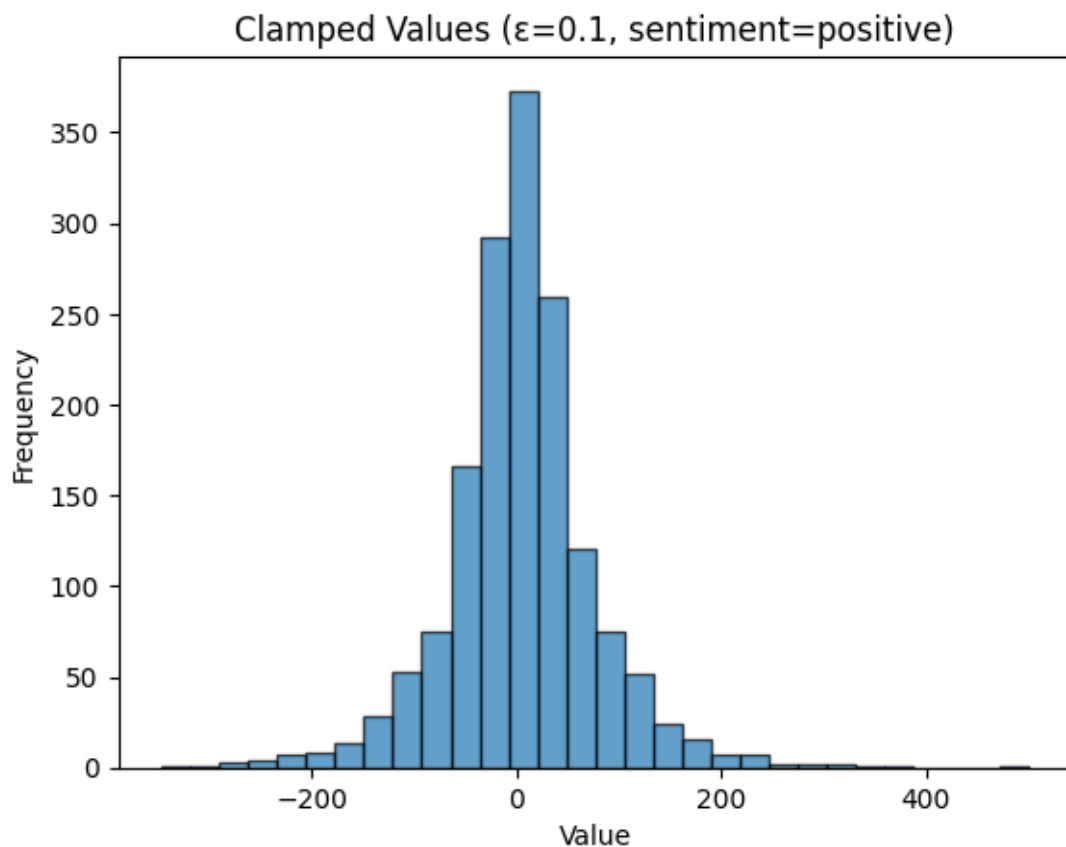
```
negative :
[(' ', 24.385287749643), (' ', 24.385287749643), (' ', 24.385287749643),
(' ', 24.385287749643), (' ', -20.614712250357), (' ',
-24.614712250357), (' ', -29.614712250357), (' ', -31.614712250357), (' ',
-45.614712250357), (' ', -46.614712250357), (' ', -93.614712250357), (' ',
-105.614712250357), (' ', -107.614712250357), (' ', -127.614712250357),
(' ', -131.614712250357), (' ', -145.614712250357), (' ',
-146.614712250357), (' ', -152.614712250357), (' ', -152.614712250357),
(' ', -169.614712250357)]
: 0.1
```





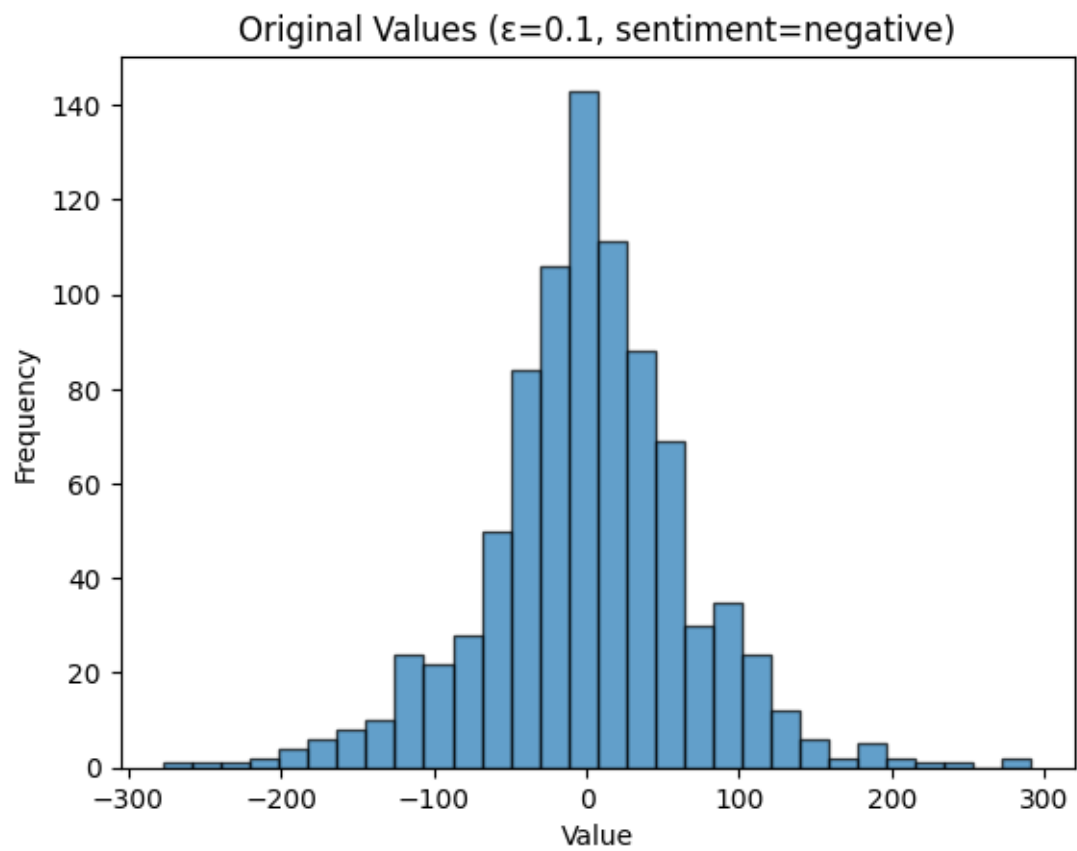
```
neutral :
[(' ', -52.41870901797978), (' ', -170.41870901797978), (' ',
-173.41870901797978), (' ', -178.41870901797978), (' ', -222.41870901797978),
(' ', -223.41870901797978), (' ', -227.41870901797978), (' ',
-229.41870901797978), (' ', -238.41870901797978), (' ',
-243.41870901797978), (' ', -256.41870901797978), (' ',
-258.41870901797978), (' ', -267.41870901797978), (' ', -277.41870901797978),
(' ', -277.41870901797978), (' ', -279.41870901797978), (' ',
-284.41870901797978), (' ', -284.41870901797978), (' ', -290.41870901797978),
(' ', -293.41870901797978)]
```

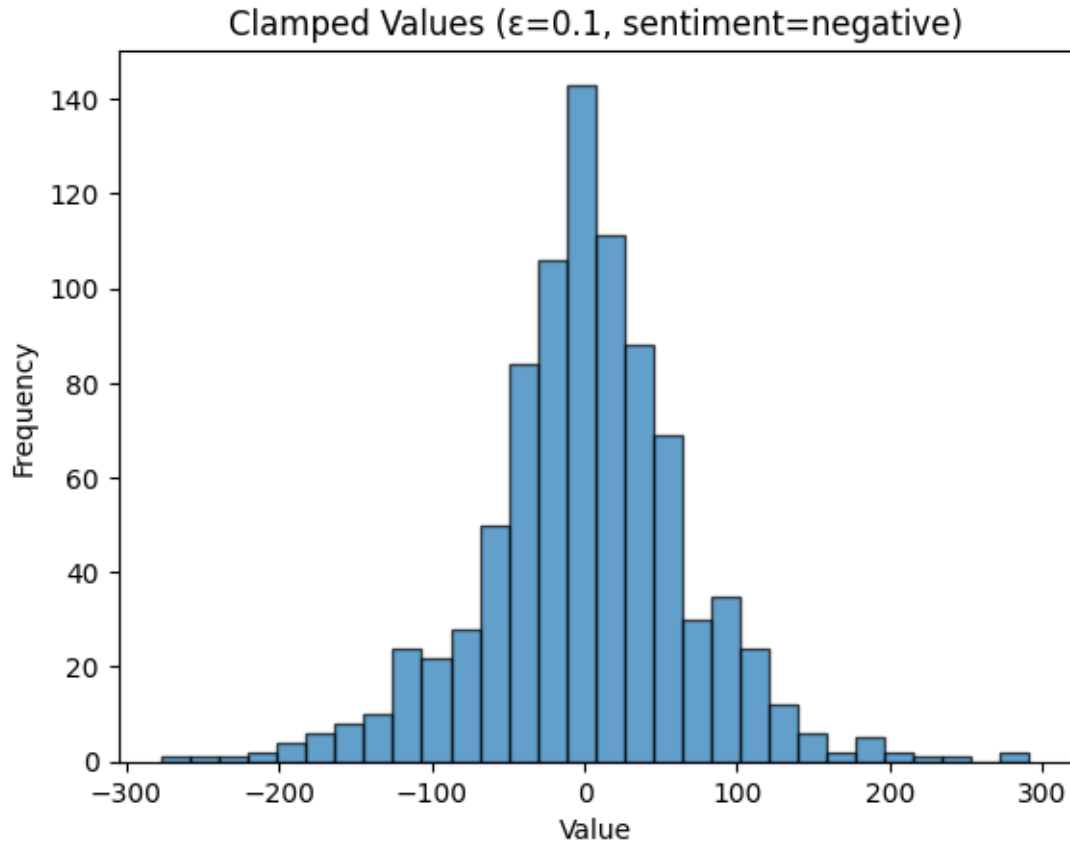




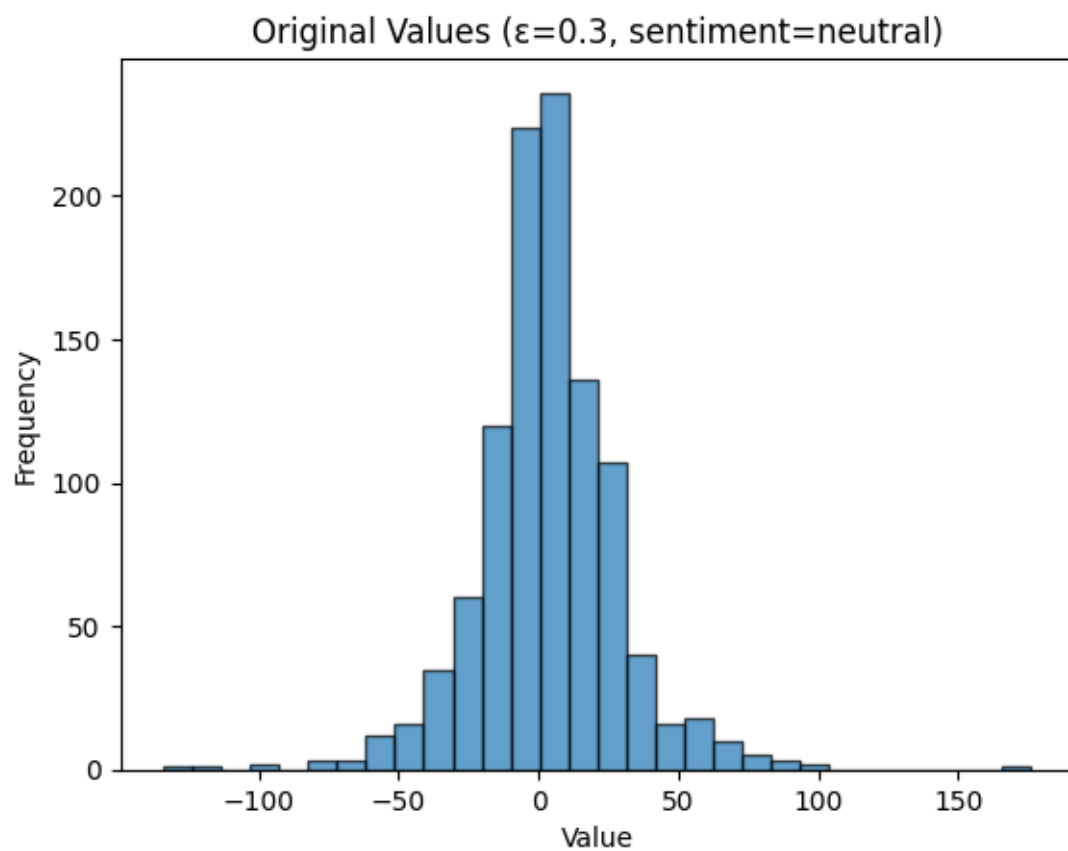
positive :

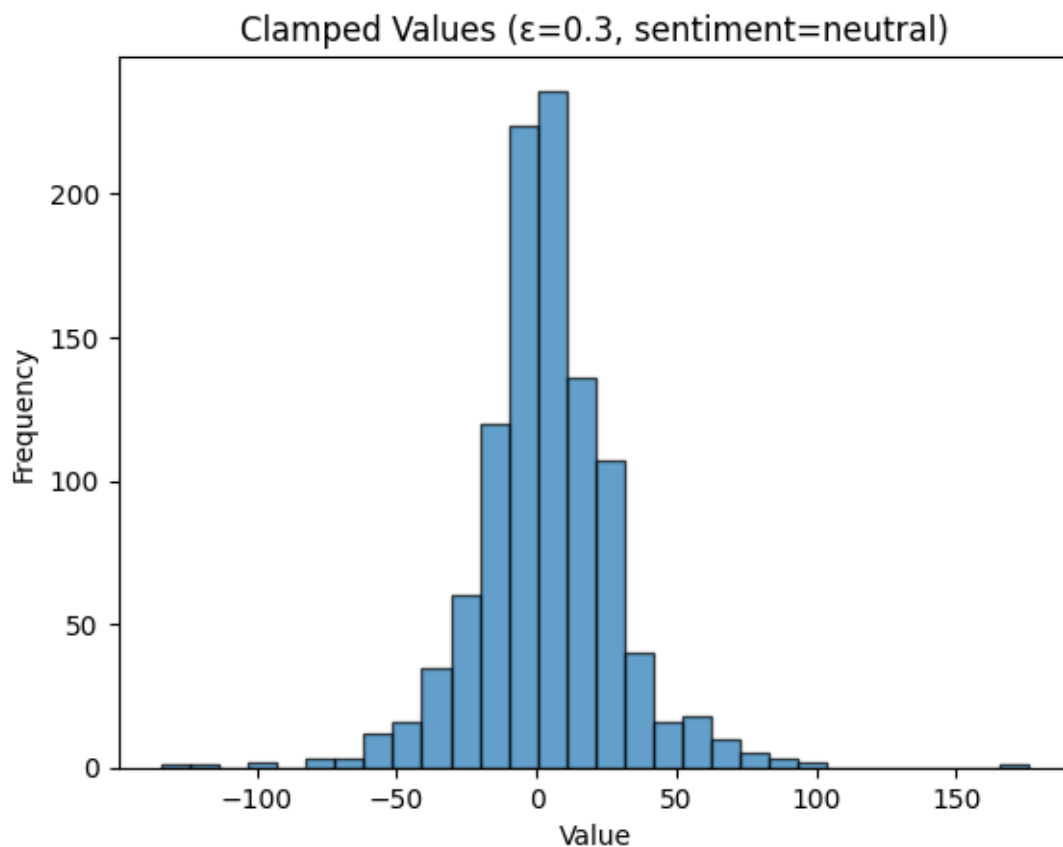
```
[(' ', 47.58129098202022), (' ', -80.41870901797978), (' ',
-121.41870901797978), (' ', -132.41870901797978), (' ', -149.41870901797978),
(' ', -156.41870901797978), (' ', -173.41870901797978), (' ',
-199.41870901797978), (' ', -203.41870901797978), (' ', -209.41870901797978),
(' ', -214.41870901797978), (' ', -220.41870901797978), (' ',
-220.41870901797978), (' ', -220.41870901797978), (' ', -221.41870901797978),
(' ', -227.41870901797978), (' ', -235.41870901797978), (' ',
-242.41870901797978), (' ', -247.41870901797978), (' ', -247.41870901797978)]
```





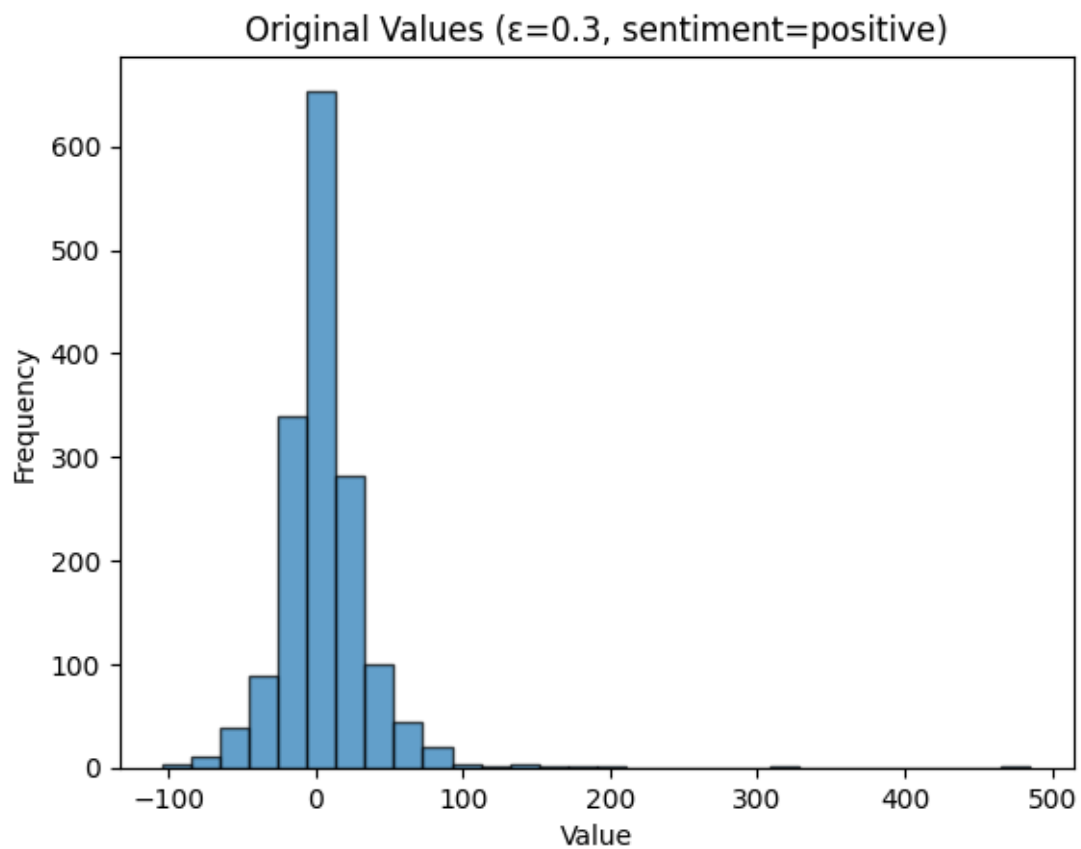
```
negative :
[(' ', -161.41870901797978), (' ', -177.41870901797978), (' ',
-213.41870901797978), (' ', -224.41870901797978), (' ', -239.41870901797978),
(' ', -252.41870901797978), (' ', -258.41870901797978), (' ',
-259.41870901797978), (' ', -260.41870901797978), (' ', -264.41870901797978),
(' ', -266.41870901797978), (' ', -275.41870901797978), (' ',
-282.41870901797978), (' ', -296.41870901797978), (' ', -297.41870901797978),
(' ', -299.41870901797978), (' ', -307.41870901797978), (' ',
-309.41870901797978), (' ', -311.41870901797978), (' ', -313.41870901797978)]
: 0.3
```

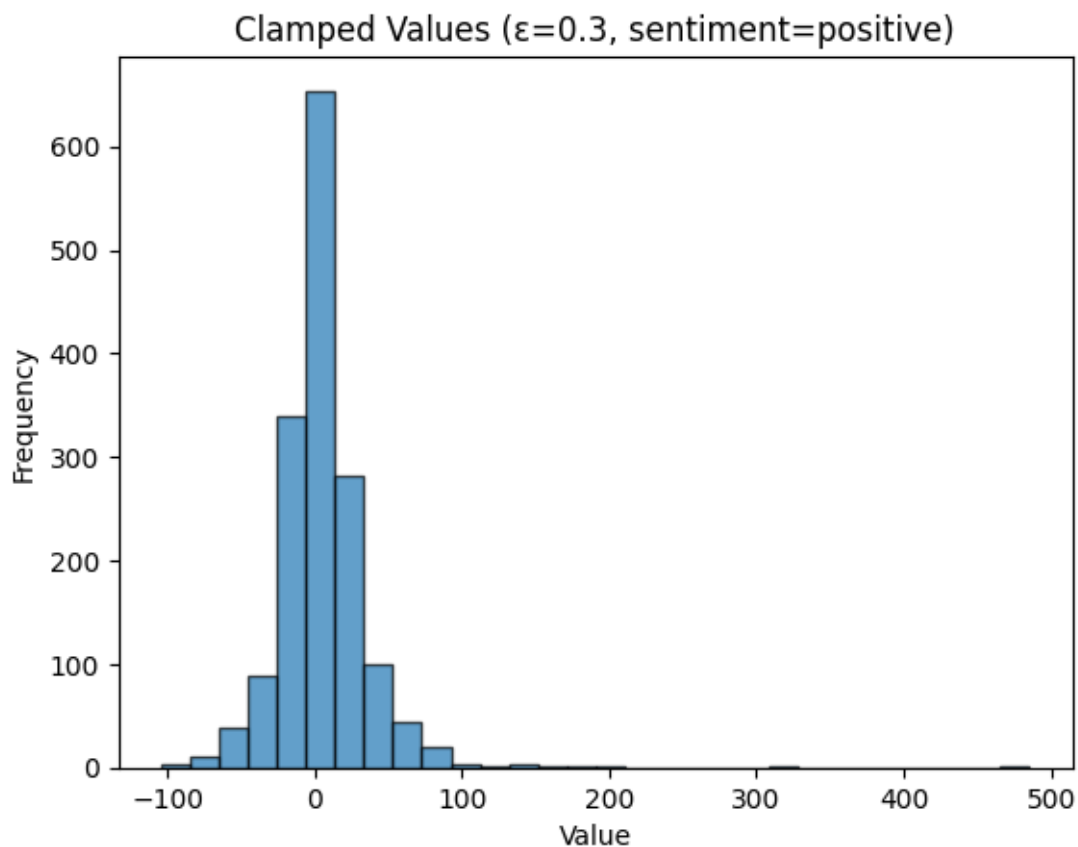





neutral :

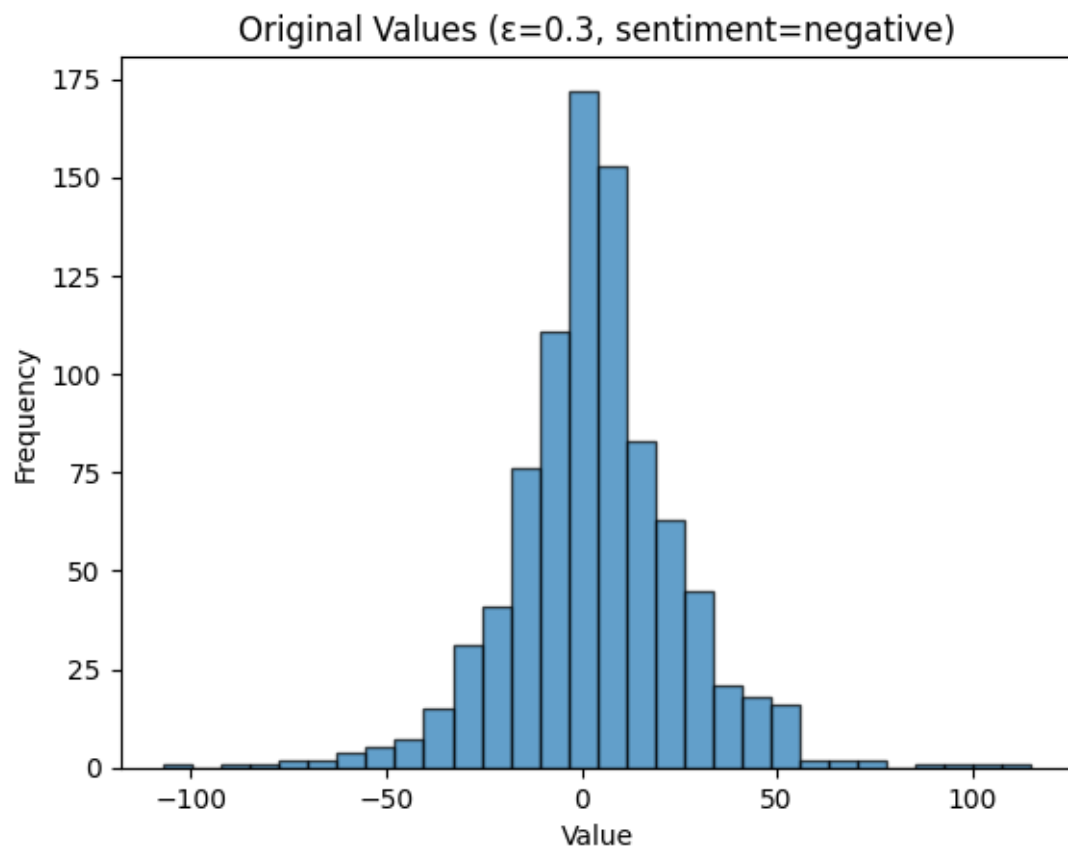
```
[(' ', -194.40911034085894), (' ', -269.40911034085894), (' ',
-275.40911034085894), (' ', -280.40911034085894), (' ', -280.40911034085894),
(' ', -283.40911034085894), (' ', -288.40911034085894), (' ',
-289.40911034085894), (' ', -290.40911034085894), (' ',
-290.40911034085894), (' ', -291.40911034085894), (' ', -298.40911034085894),
(' ', -298.40911034085894), (' ', -299.40911034085894), (' ',
-300.40911034085894), (' ', -301.40911034085894), (' ', -301.40911034085894),
(' ', -302.40911034085894), (' ', -302.40911034085894), (' ',
-303.40911034085894)]
```

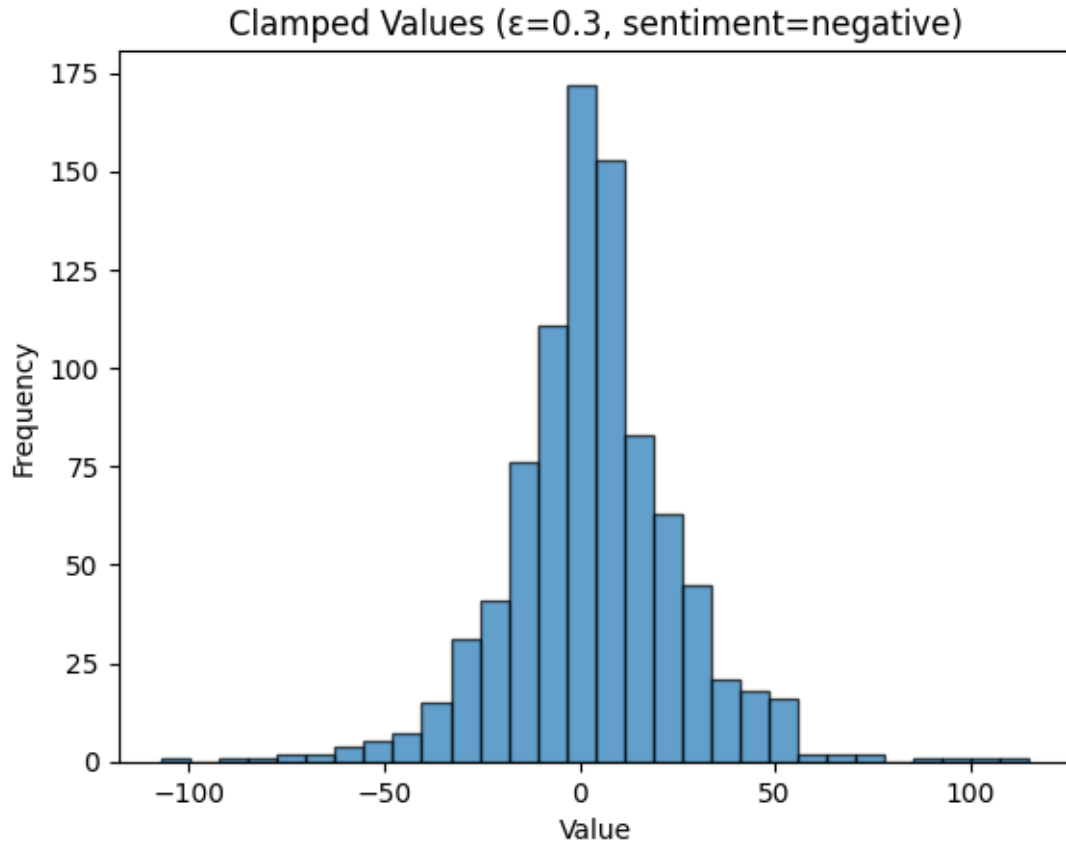




positive :

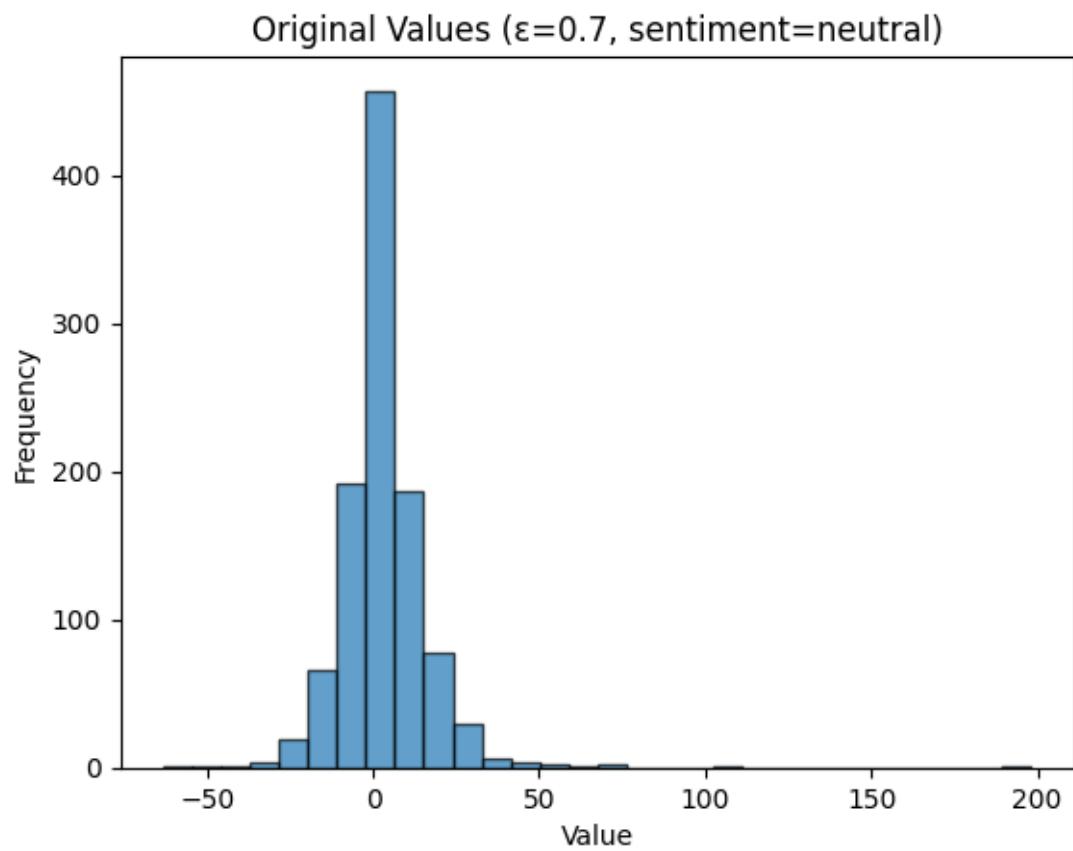
```
[(' ', 114.59088965914106), (' ', -54.409110340858945), (' ',
-160.40911034085894), (' ', -172.40911034085894), (' ', -190.40911034085894),
(' ', -194.40911034085894), (' ', -207.40911034085894), (' ',
-212.40911034085894), (' ', -231.40911034085894), (' ', -235.40911034085894),
(' ', -238.40911034085894), (' ', -254.40911034085894), (' ',
-266.40911034085894), (' ', -269.40911034085894), (' ', -275.40911034085894),
(' ', -277.40911034085894), (' ', -279.40911034085894), (' ',
-280.40911034085894), (' ', -280.40911034085894), (' ',
-283.40911034085894)]
```

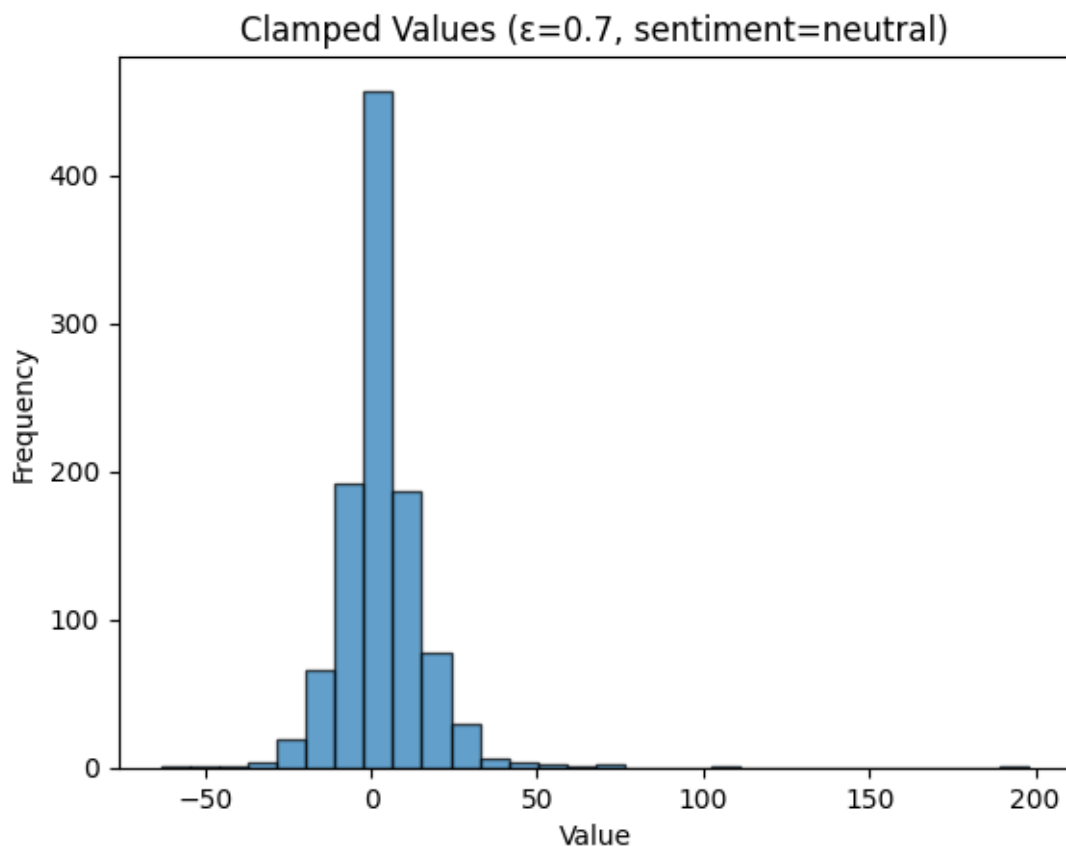




negative :

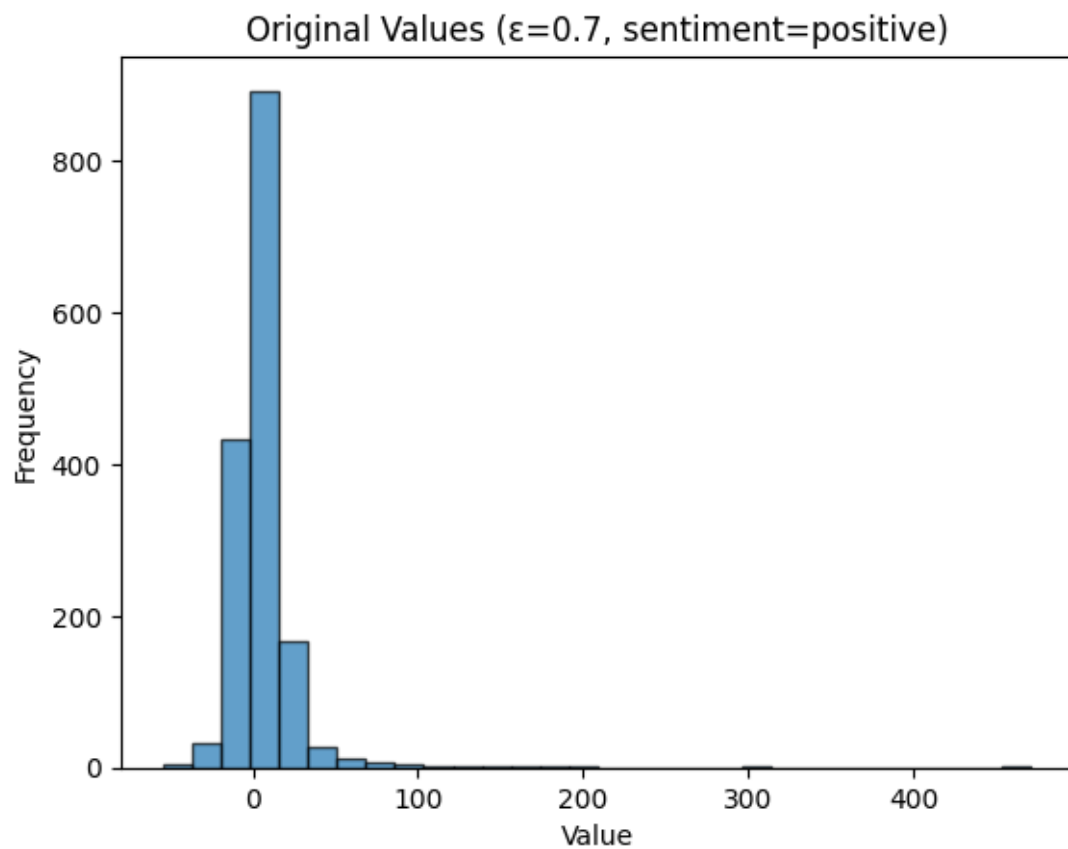
```
[(' ', -255.40911034085894), (' ', -264.40911034085894), (' ',
-271.40911034085894), (' ', -279.40911034085894), (' ',
-293.40911034085894), (' ', -297.40911034085894), (' ', -300.40911034085894),
(' ', -301.40911034085894), (' ', -308.40911034085894), (' ',
-313.40911034085894), (' ', -315.40911034085894), (' ',
-316.40911034085894), (' ', -316.40911034085894), (' ', -316.40911034085894),
(' ', -316.40911034085894), (' ', -317.40911034085894), (' ',
-318.40911034085894), (' ', -319.40911034085894), (' ',
-319.40911034085894), (' ', -320.40911034085894)]
: 0.7
```

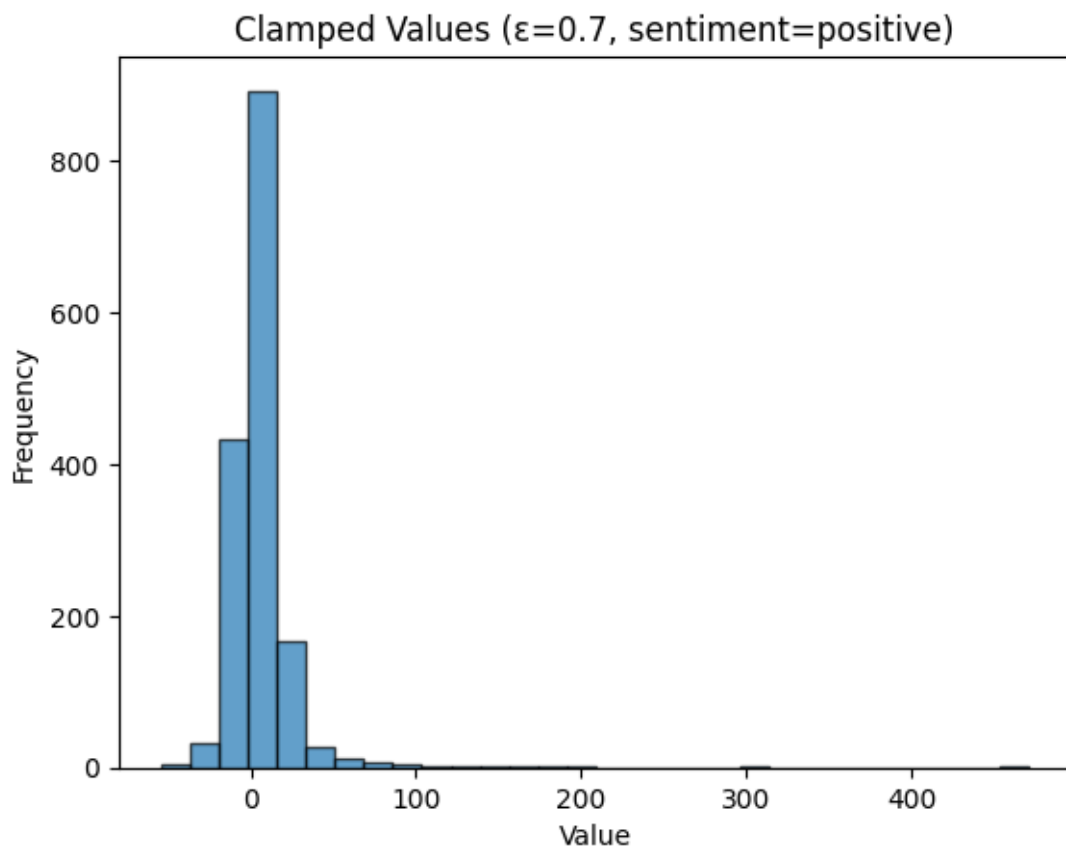




neutral :

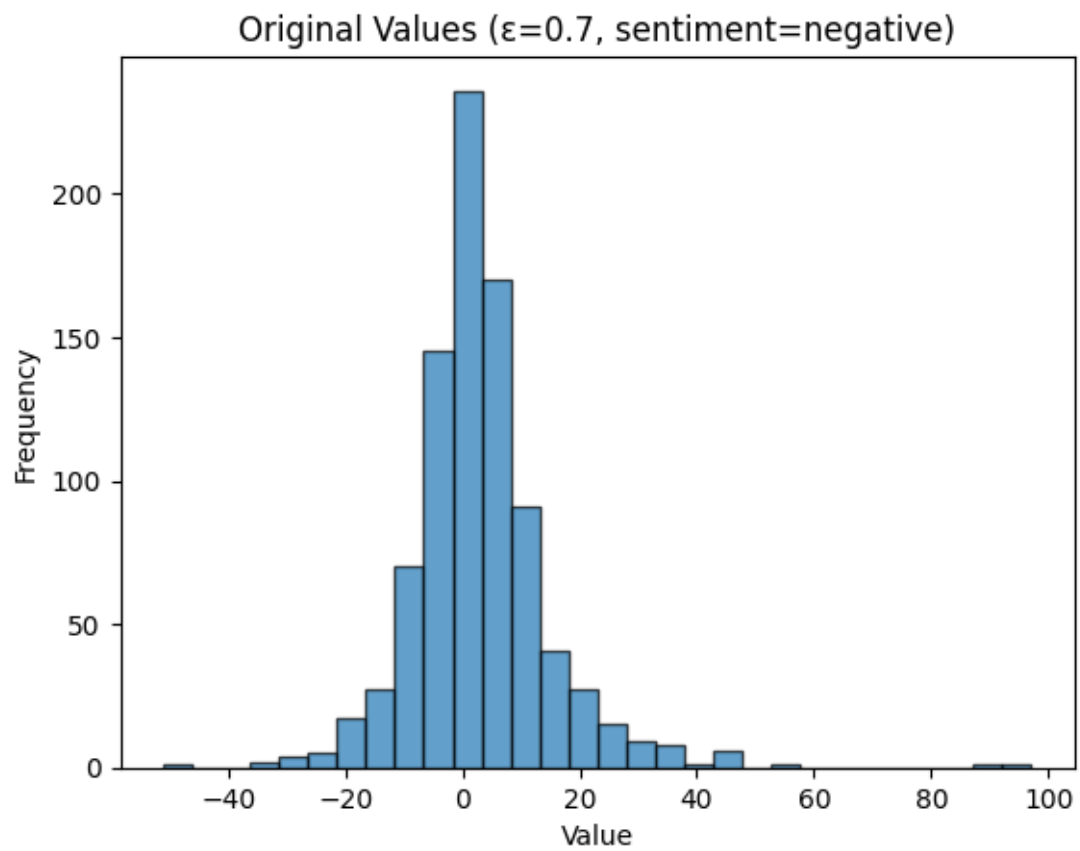
```
[(' ', -50.29265189570475), (' ', -141.29265189570475), (' ',
-178.29265189570475), (' ', -179.29265189570475), (' ', -187.29265189570475),
(' ', -190.29265189570475), (' ', -195.29265189570475), (' ',
-198.29265189570475), (' ', -203.29265189570475), (' ', -205.29265189570475),
(' ', -208.29265189570475), (' ', -209.29265189570475), (' ',
-211.29265189570475), (' ', -213.29265189570475), (' ',
-213.29265189570475), (' ', -215.29265189570475), (' ', -216.29265189570475),
(' ', -217.29265189570475), (' ', -217.29265189570475), (' ',
-217.29265189570475)]
```

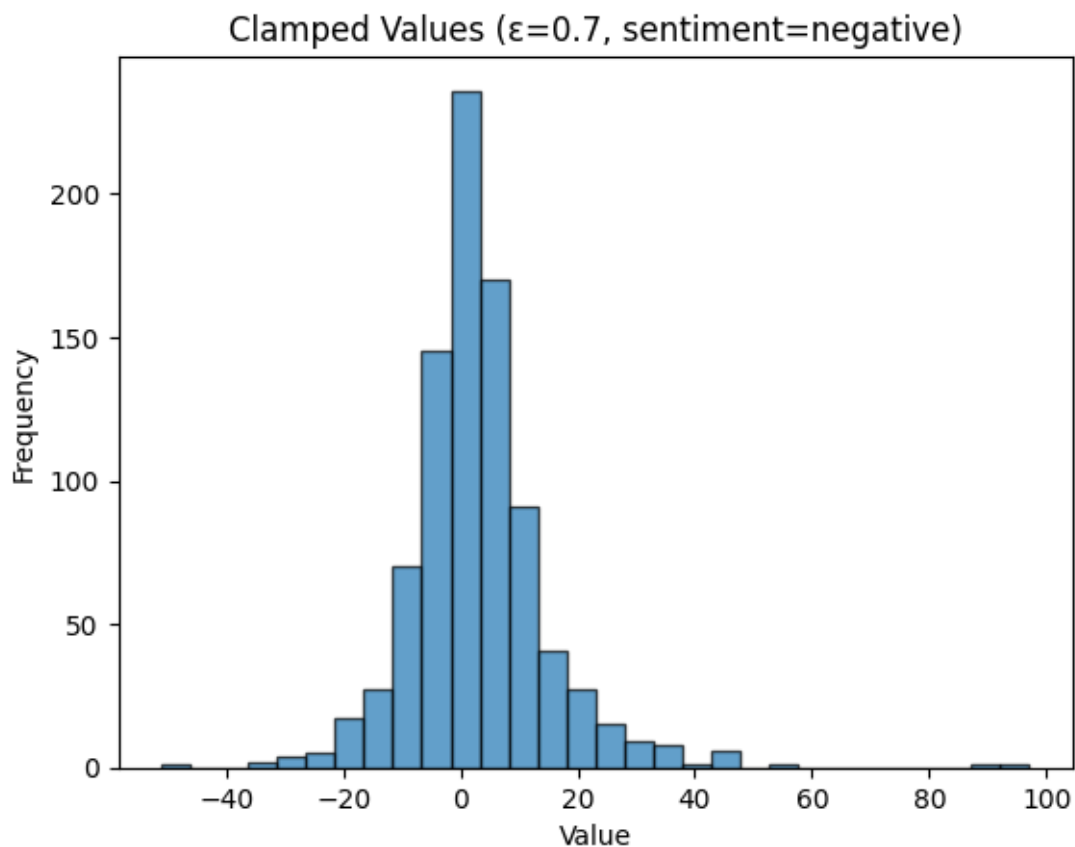





positive :

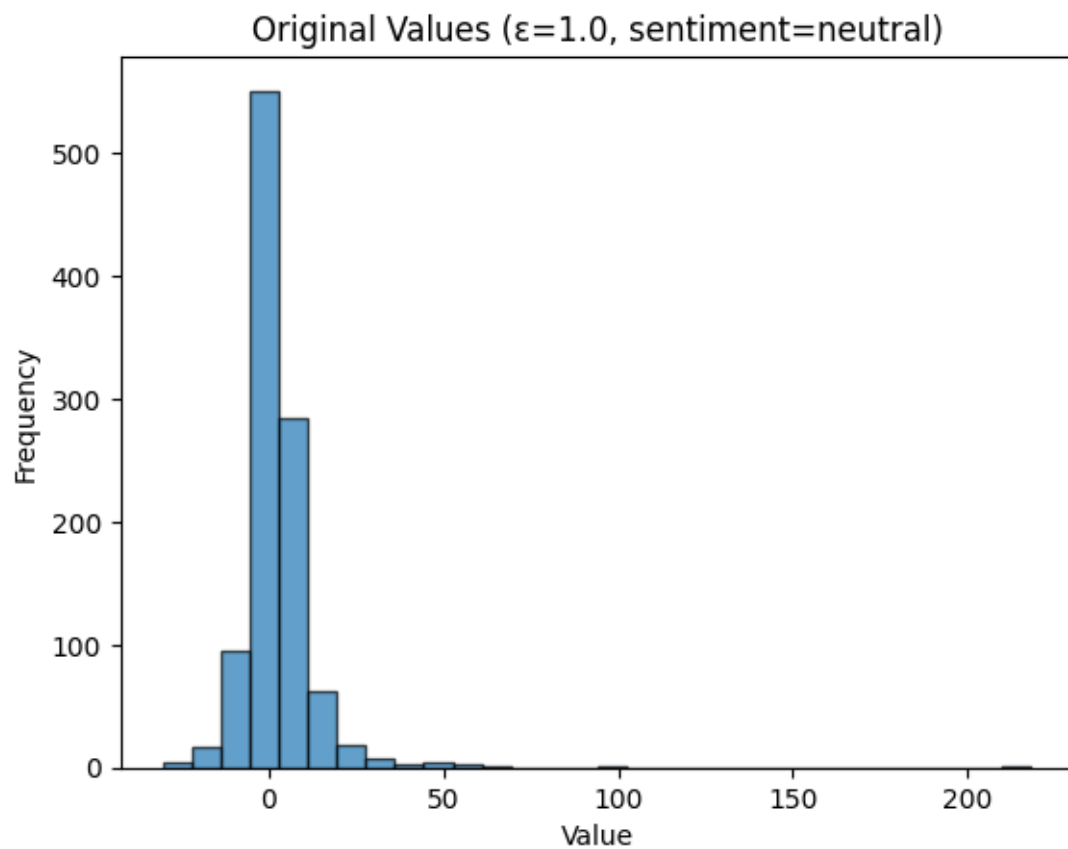
```
[(' ', 222.70734810429525), (' ', 63.70734810429525), (' ',
-42.29265189570475), (' ', -54.29265189570475), (' ', -64.29265189570475),
(' ', -69.29265189570475), (' ', -75.29265189570475), (' ',
-86.29265189570475), (' ', -98.29265189570475), (' ', -108.29265189570475),
(' ', -122.29265189570475), (' ', -137.29265189570475), (' ',
-142.29265189570475), (' ', -142.29265189570475), (' ', -149.29265189570475),
(' ', -150.29265189570475), (' ', -151.29265189570475), (' ',
-161.29265189570475), (' ', -163.29265189570475), (' ', -165.29265189570475)]
```

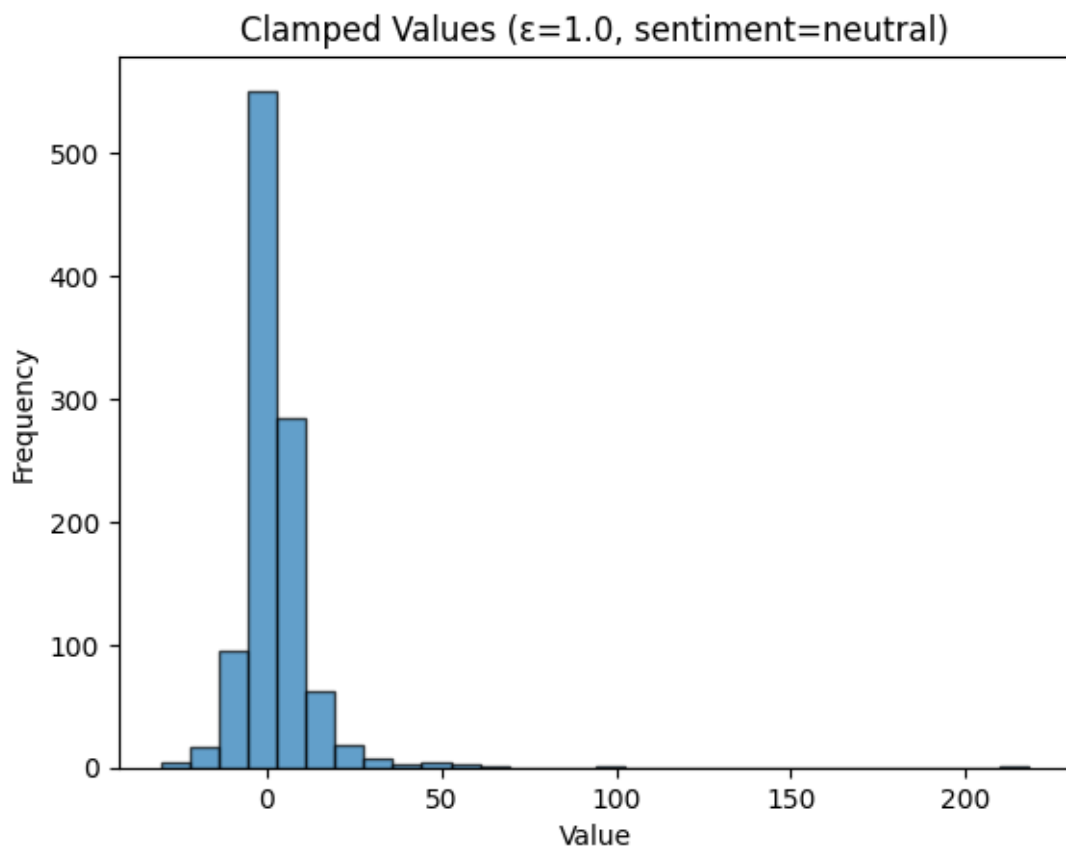




negative :

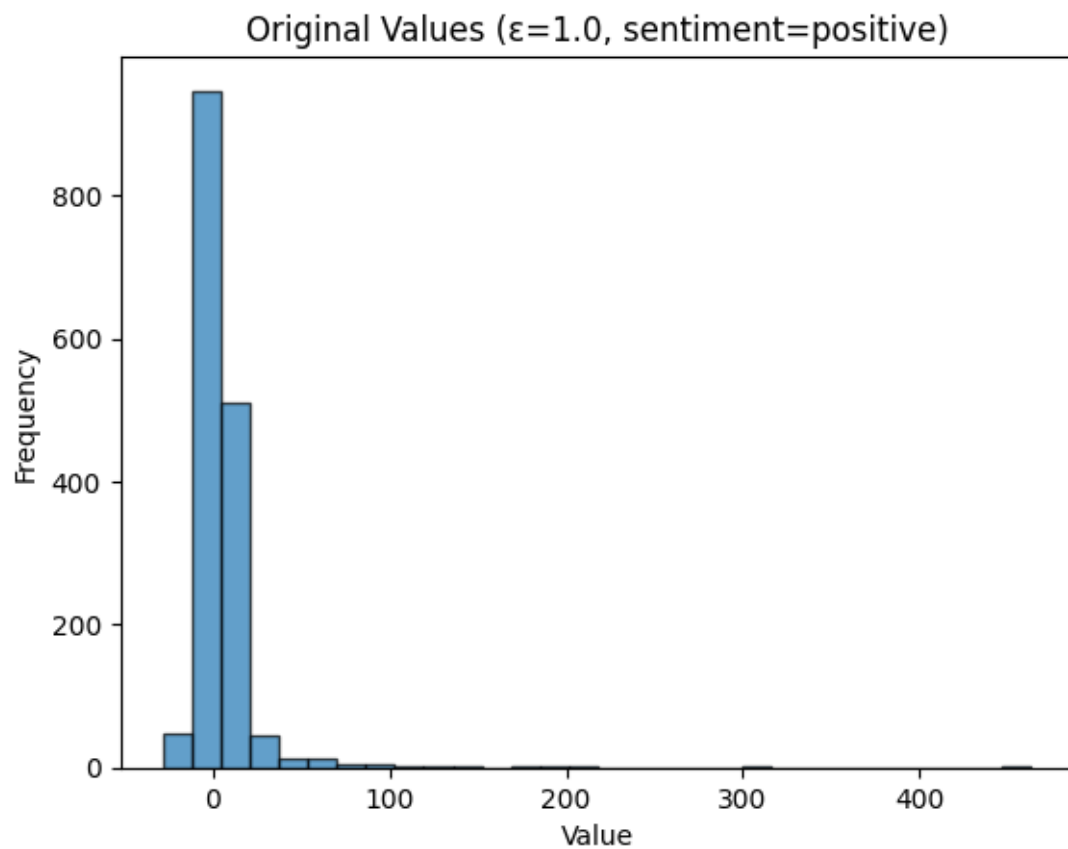
```
[(' ', -151.29265189570475), (' ', -158.29265189570475), (' ',
-192.29265189570475), (' ', -202.29265189570475), (' ', -203.29265189570475),
(' ', -203.29265189570475), (' ', -204.29265189570475), (' ',
-204.29265189570475), (' ', -205.29265189570475), (' ', -209.29265189570475),
(' ', -211.29265189570475), (' ', -211.29265189570475), (' ',
-212.29265189570475), (' ', -213.29265189570475), (' ', -213.29265189570475),
(' ', -214.29265189570475), (' ', -214.29265189570475), (' ',
-214.29265189570475), (' ', -216.29265189570475), (' ', -217.29265189570475)]
: 1.0
```

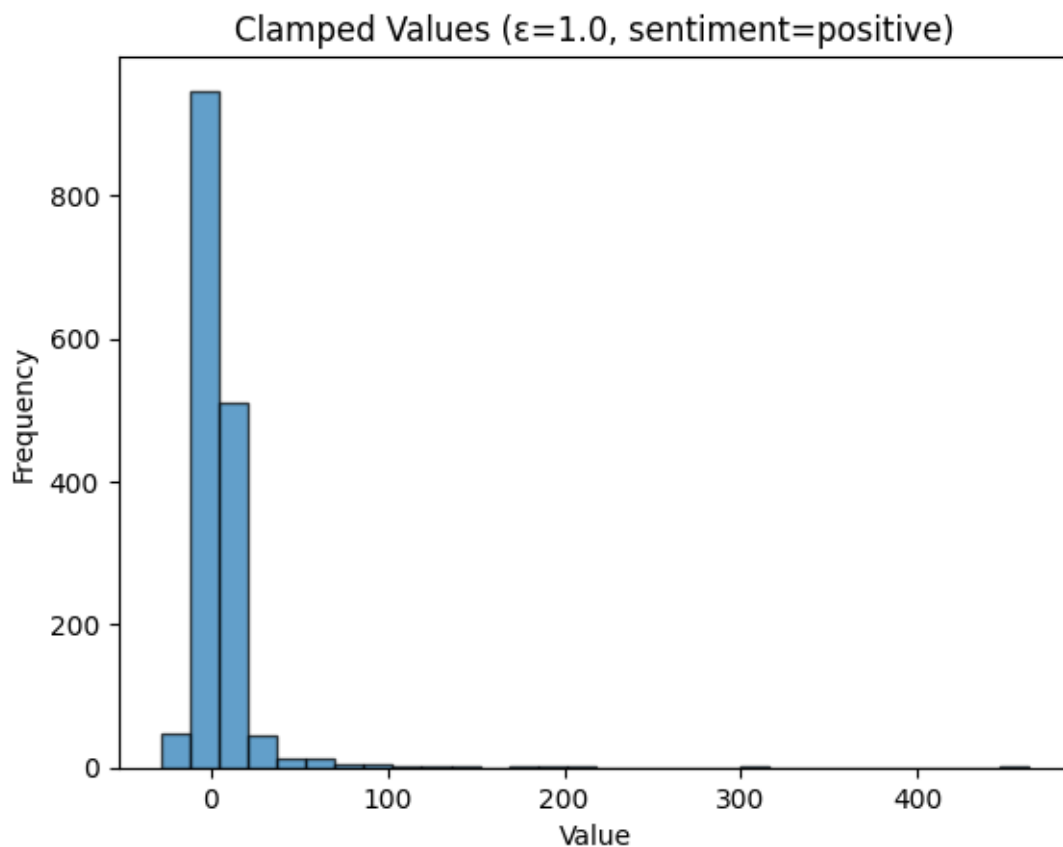




neutral :

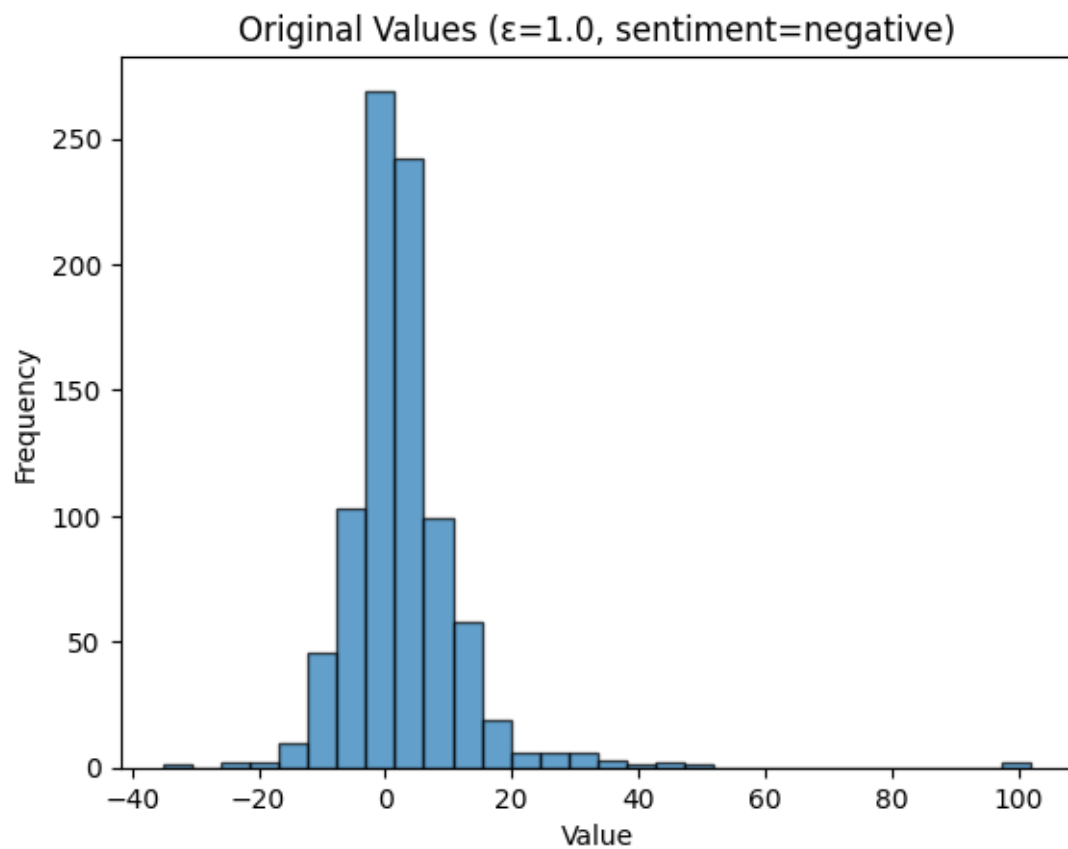
```
[(' ', 34.060279414278824), (' ', -86.93972058572118), (' ',
-116.93972058572118), (' ', -126.93972058572118), (' ', -126.93972058572118),
(' ', -132.93972058572118), (' ', -134.93972058572118), (' ',
-135.93972058572118), (' ', -138.93972058572118), (' ', -140.93972058572118),
(' ', -141.93972058572118), (' ', -146.93972058572118), (' ',
-151.93972058572118), (' ', -151.93972058572118), (' ', -151.93972058572118),
(' ', -153.93972058572118), (' ', -153.93972058572118), (' ',
-154.93972058572118), (' ', -155.93972058572118), (' ', -156.93972058572118)]
```

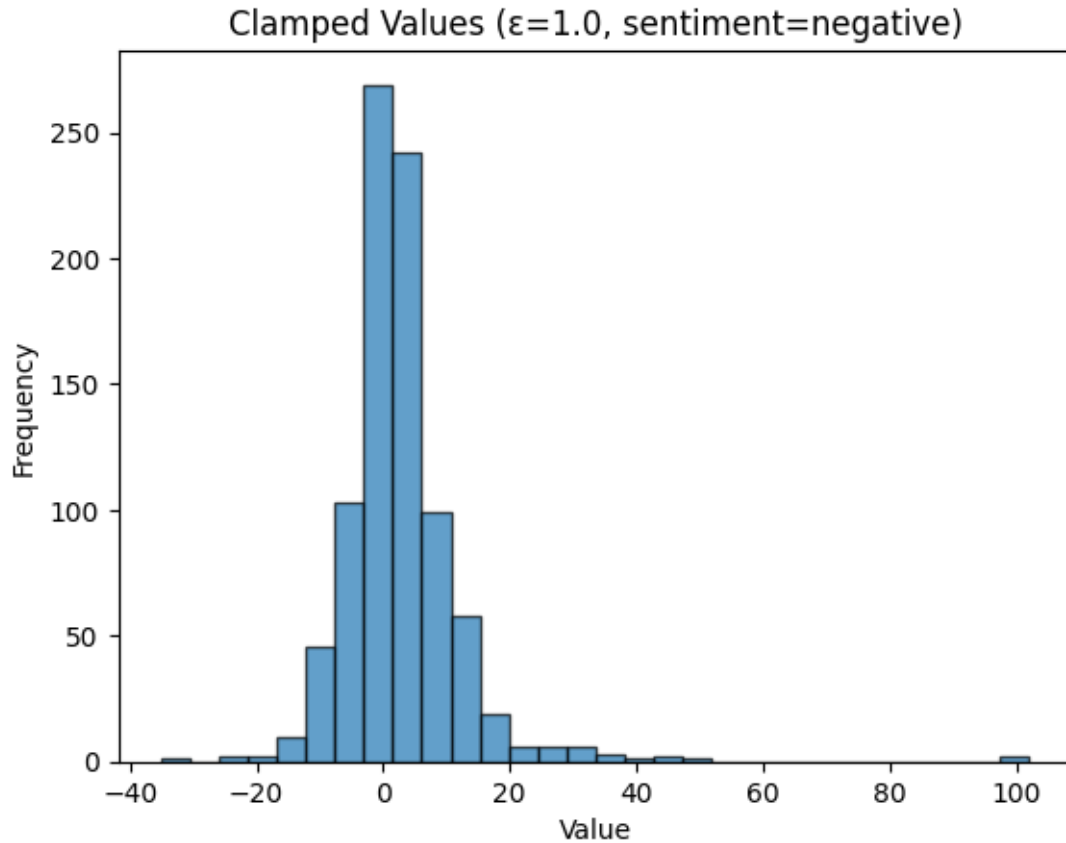




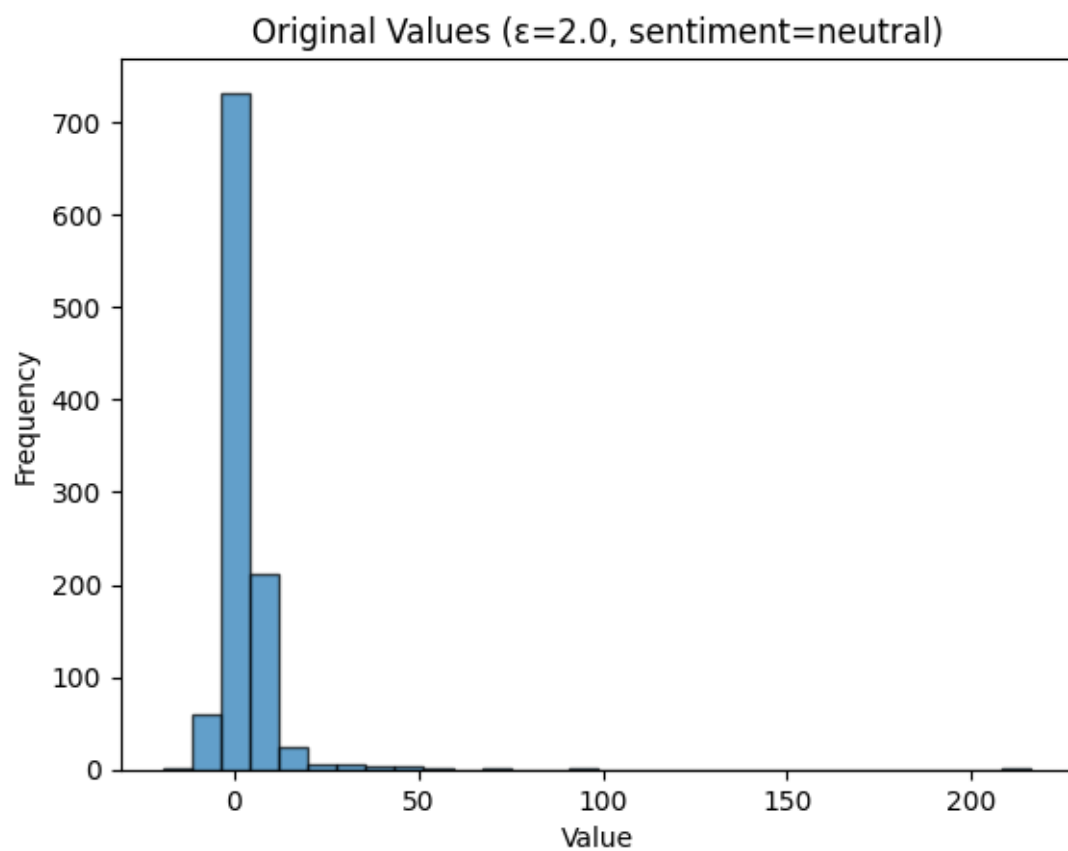
positive :

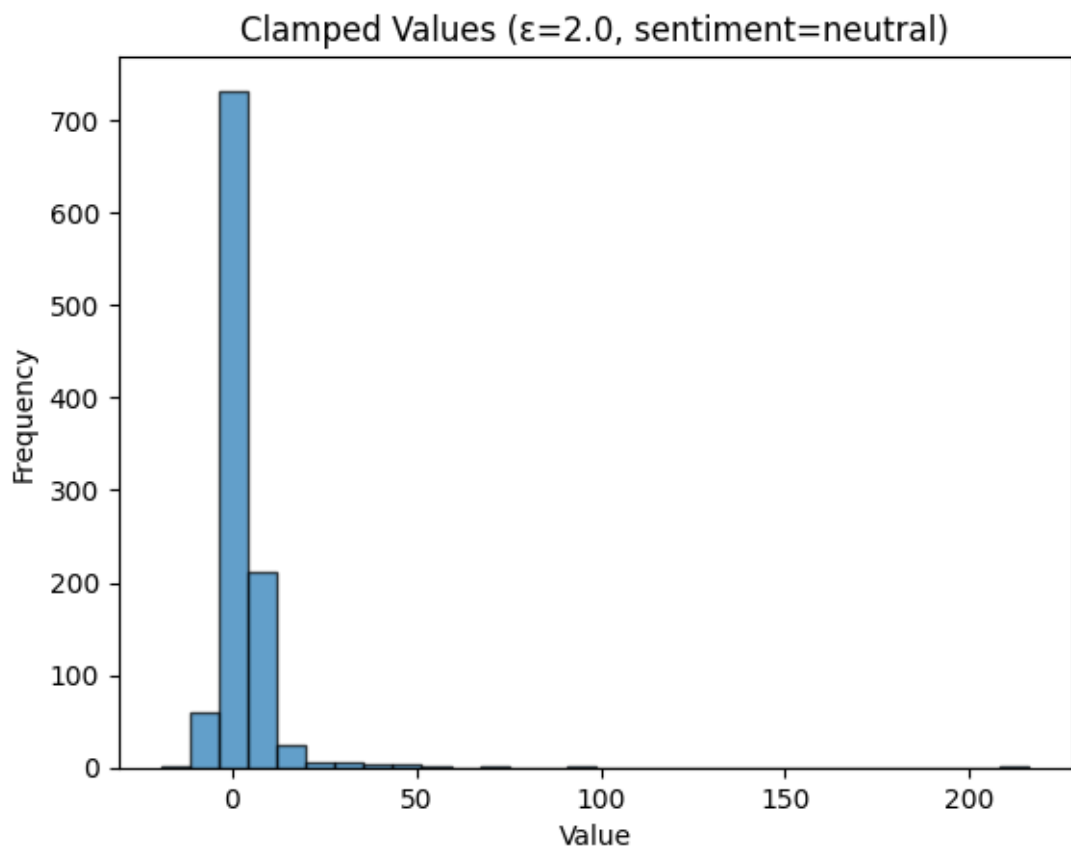
```
[(' ', 279.06027941427885), (' ', 122.06027941427882), (' ',
32.060279414278824), (' ', 24.060279414278824), (' ', 9.060279414278824),
(' ', -2.939720585721176), (' ', -10.939720585721176), (' ',
-41.939720585721176), (' ', -43.939720585721176), (' ', -49.939720585721176),
(' ', -54.939720585721176), (' ', -75.93972058572118), (' ',
-78.93972058572118), (' ', -84.93972058572118), (' ', -87.93972058572118),
(' ', -92.93972058572118), (' ', -92.93972058572118), (' ',
-95.93972058572118), (' ', -97.93972058572118), (' ', -102.93972058572118)]
```



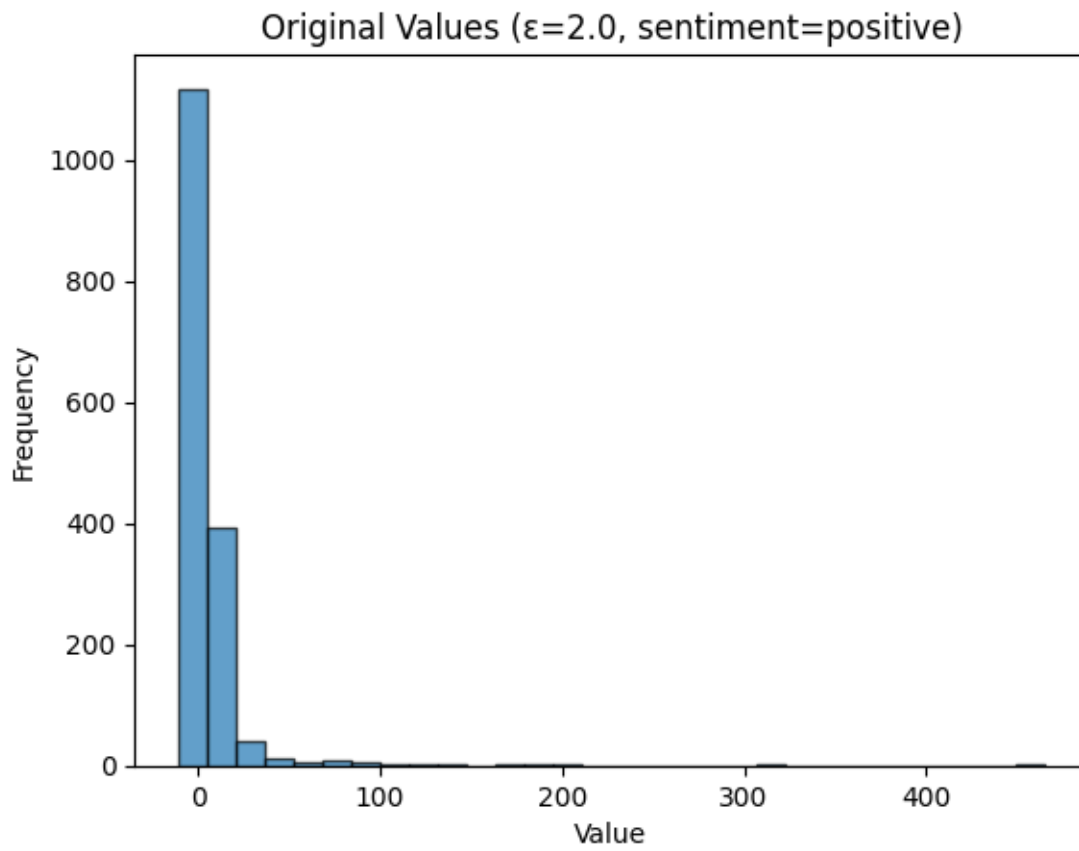
```
negative :
[(' ', -81.93972058572118), (' ', -82.93972058572118), (' ',
-133.93972058572118), (' ', -138.93972058572118), (' ', -139.93972058572118),
(' ', -144.93972058572118), (' ', -145.93972058572118), (' ',
-146.93972058572118), (' ', -149.93972058572118), (' ',
-151.93972058572118), (' ', -151.93972058572118), (' ', -151.93972058572118),
(' ', -152.93972058572118), (' ', -153.93972058572118), (' ',
-154.93972058572118), (' ', -155.93972058572118), (' ', -155.93972058572118),
(' ', -157.93972058572118), (' ', -157.93972058572118), (' ',
-158.93972058572118)]
: 2.0
```

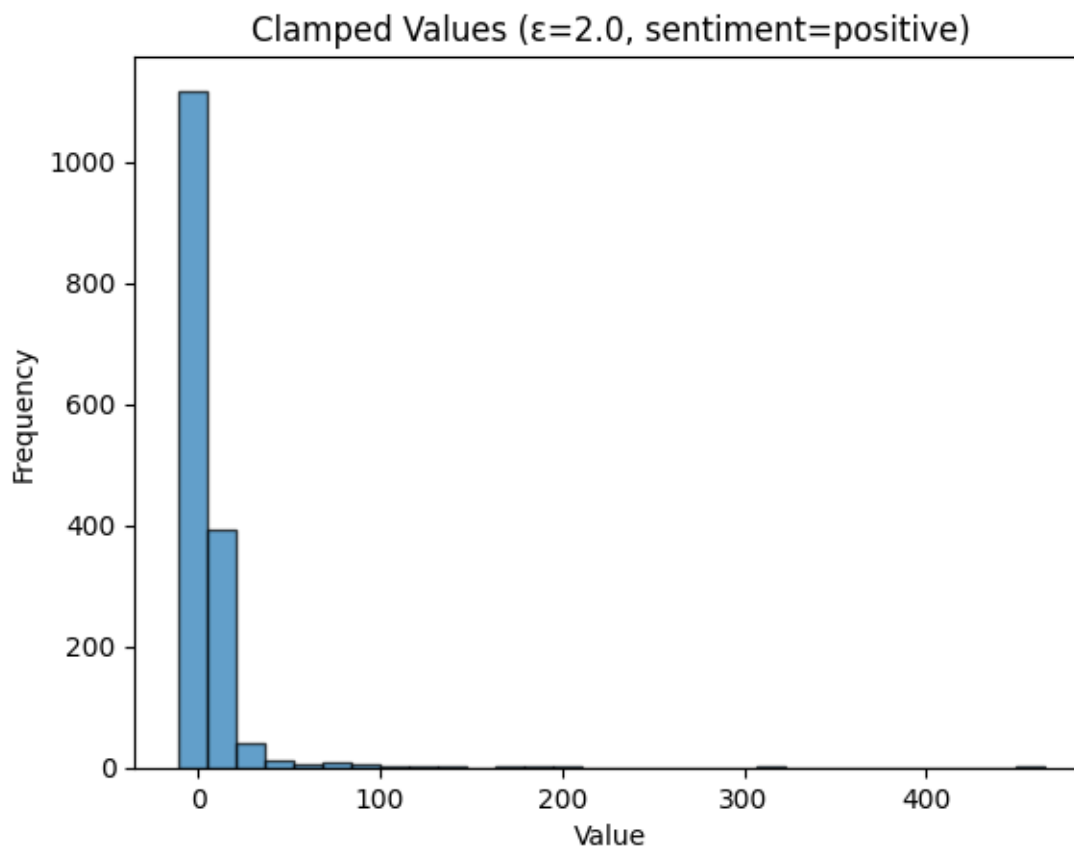




neutral :

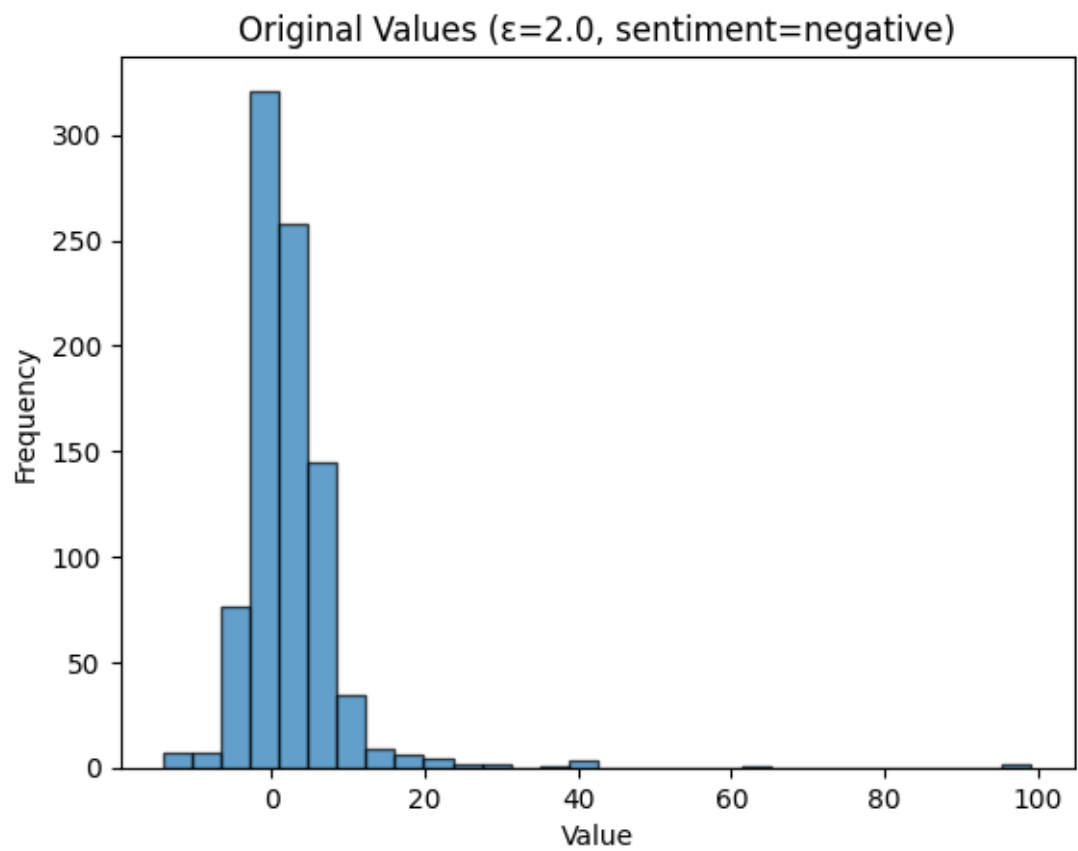
```
[(' ', 148.33235838169367), (' ', 23.332358381693652), (' ',
0.33235838169365195), (' ', -9.667641618306348), (' ', -12.667641618306348),
(' ', -17.667641618306348), (' ', -19.667641618306348), (' ',
-23.667641618306348), (' ', -24.667641618306348), (' ', -31.667641618306348),
(' ', -31.667641618306348), (' ', -34.66764161830635), (' ',
-37.66764161830635), (' ', -38.66764161830635), (' ', -38.66764161830635),
(' ', -39.66764161830635), (' ', -41.66764161830635), (' ',
-42.66764161830635), (' ', -43.66764161830635), (' ', -43.66764161830635)]
```

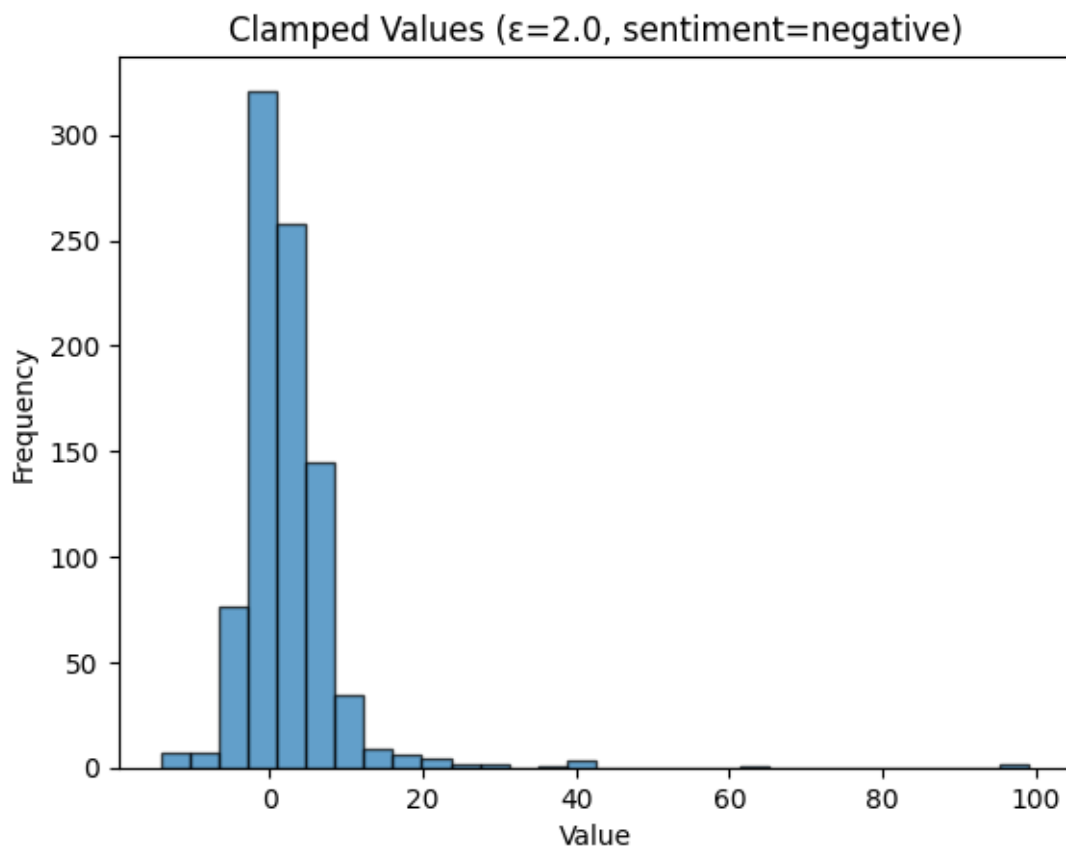




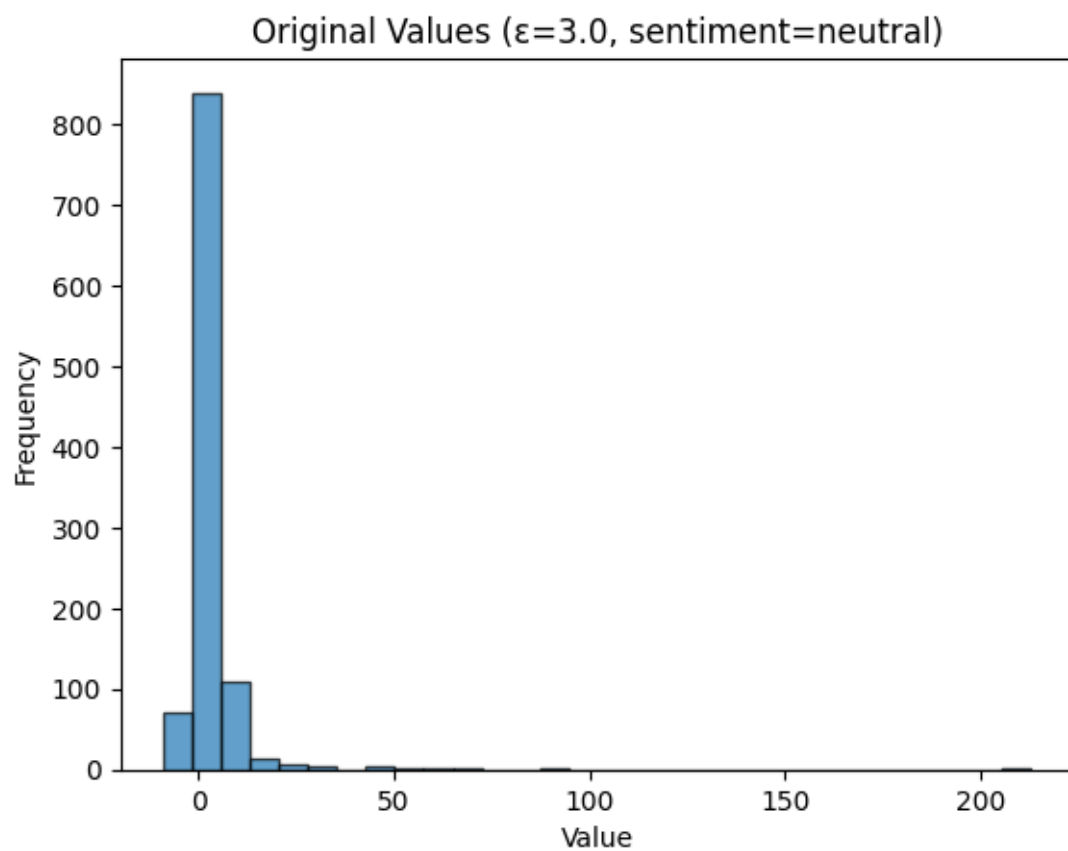
positive :

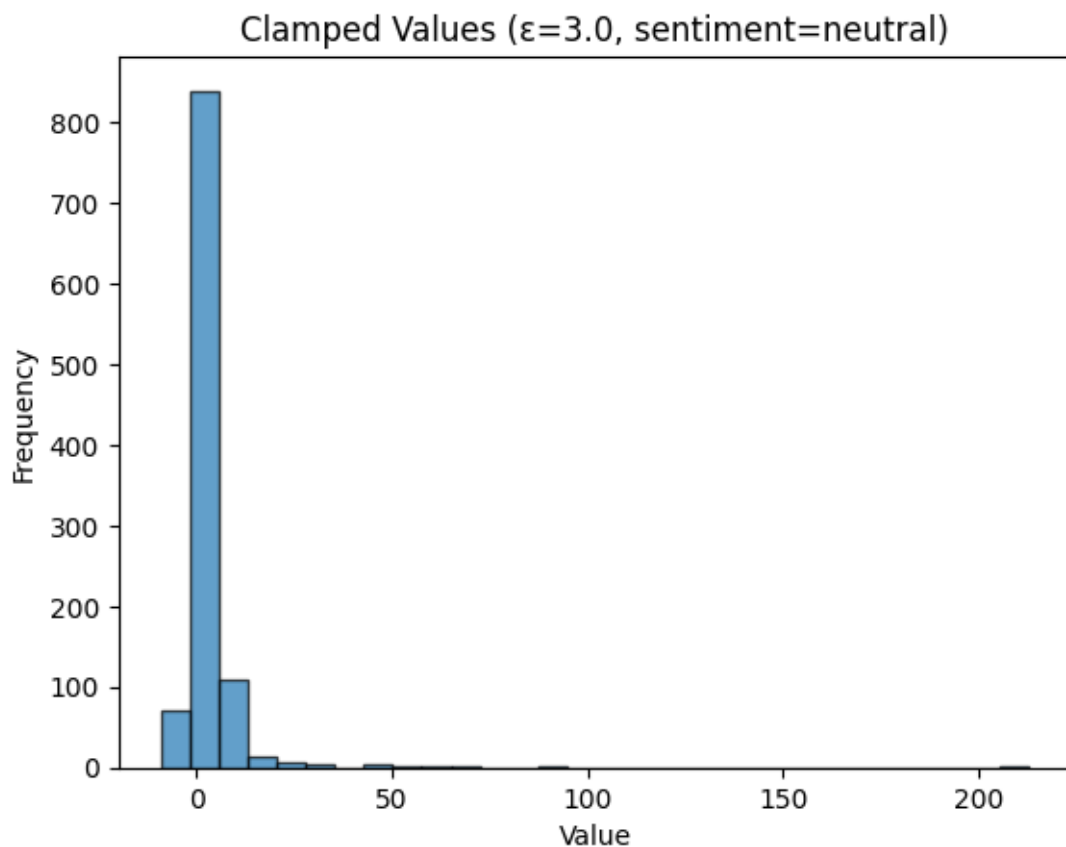
```
[(' ', 397.33235838169367), (' ', 250.33235838169367), (' ',
138.33235838169367), (' ', 130.33235838169367), (' ', 124.33235838169365),
(' ', 119.33235838169365), (' ', 110.33235838169365), (' ',
79.33235838169365), (' ', 72.33235838169365), (' ', 66.33235838169365),
(' ', 53.33235838169365), (' ', 40.33235838169365), (' ',
35.33235838169365), (' ', 31.332358381693652), (' ', 29.332358381693652),
(' ', 21.332358381693652), (' ', 20.332358381693652), (' ',
19.332358381693652), (' ', 18.332358381693652), (' ', 12.332358381693652)]
```





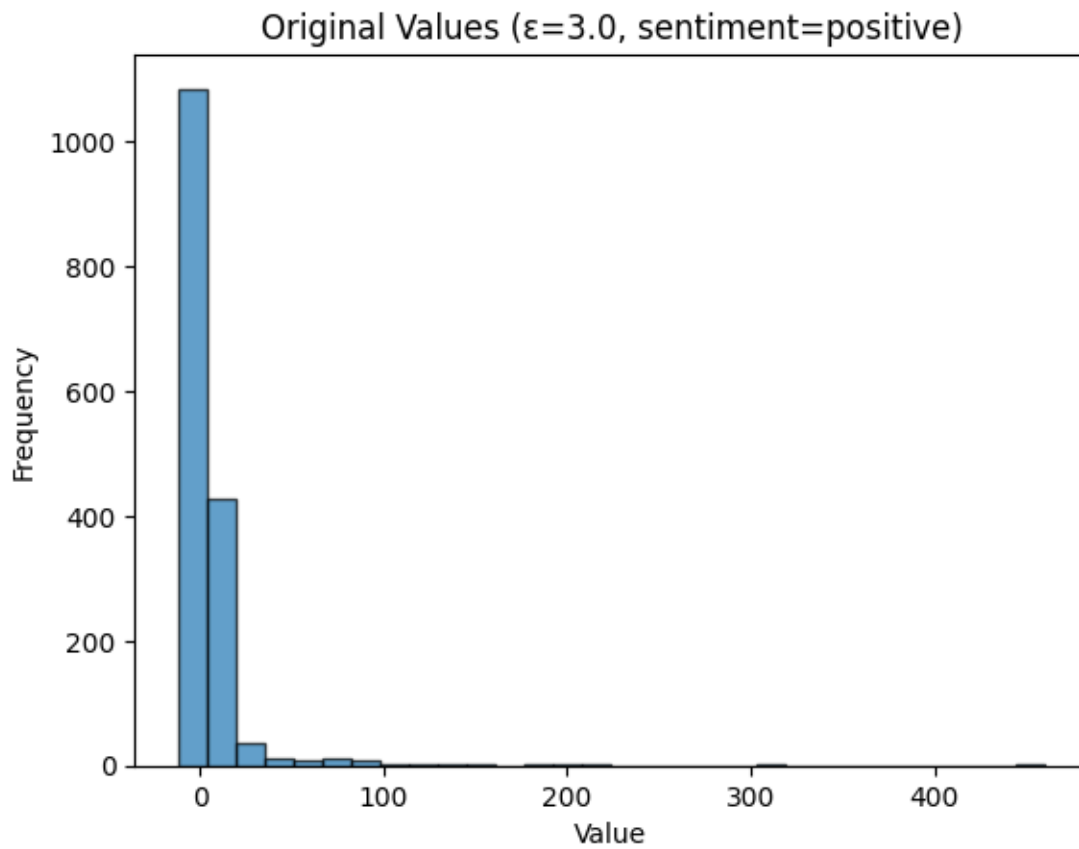
```
negative :
[(' ', 31.332358381693652), (' ', 29.332358381693652), (' ',
-4.667641618306348), (' ', -25.667641618306348), (' ', -25.667641618306348),
(' ', -26.667641618306348), (' ', -32.66764161830635), (' ',
-36.66764161830635), (' ', -38.66764161830635), (' ', -42.66764161830635),
(' ', -42.66764161830635), (' ', -44.66764161830635), (' ',
-45.66764161830635), (' ', -46.66764161830635), (' ', -47.66764161830635),
(' ', -48.66764161830635), (' ', -48.66764161830635), (' ',
-49.66764161830635), (' ', -50.66764161830635), (' ', -50.66764161830635)]
: 3.0
```

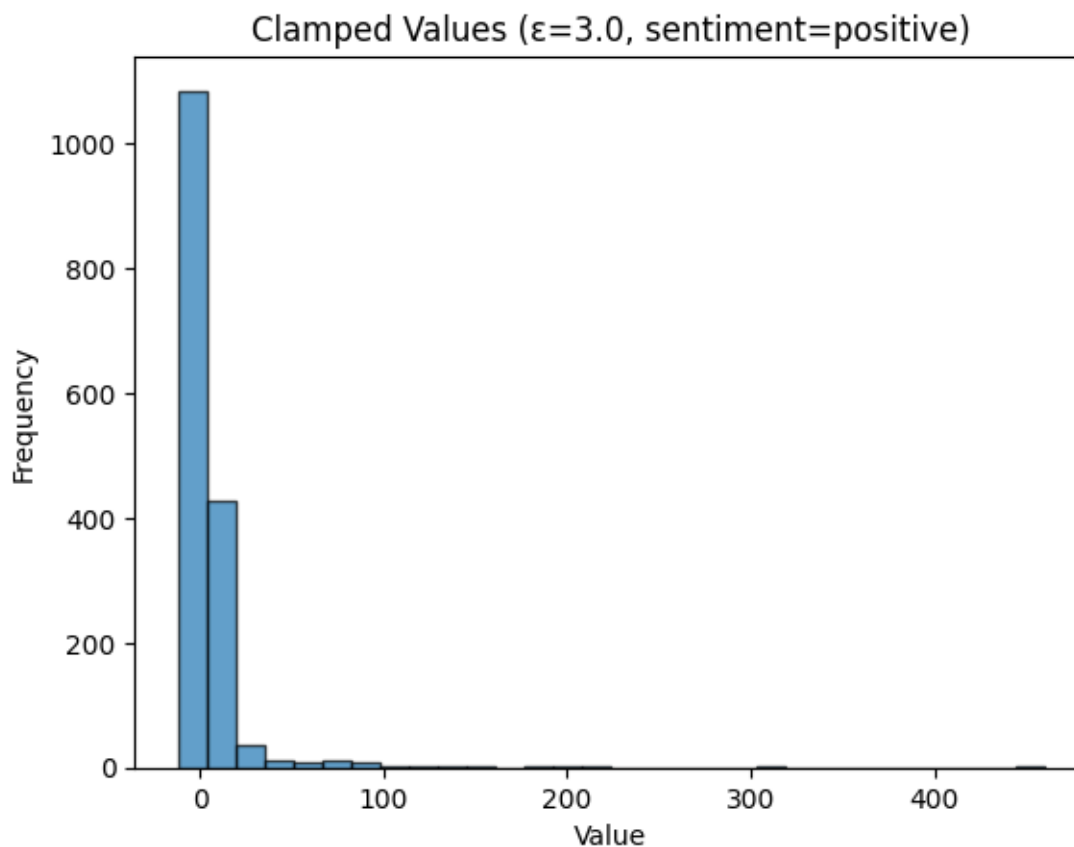





neutral :

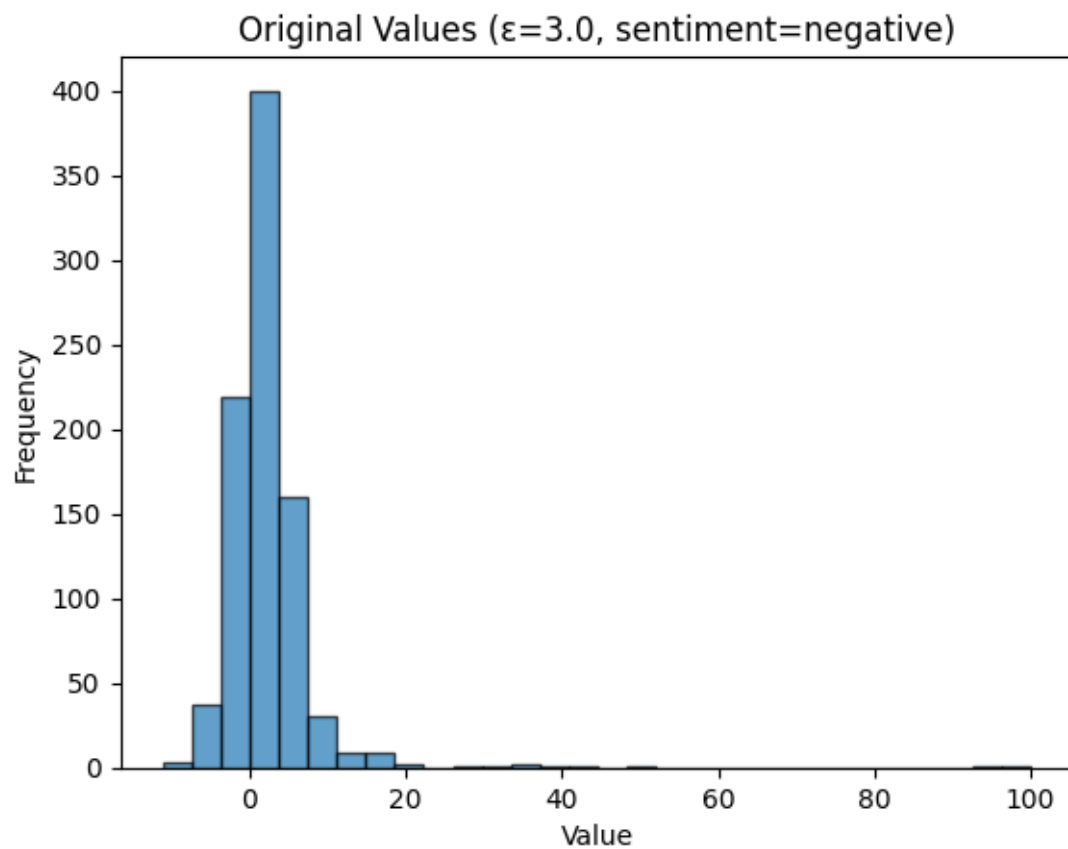
```
[(' ', 188.10646581606804), (' ', 69.10646581606802), (' ',
42.106465816068024), (' ', 35.106465816068024), (' ', 33.106465816068024),
(' ', 26.106465816068027), (' ', 26.106465816068027), (' ',
20.106465816068027), (' ', 18.106465816068027), (' ', 18.106465816068027),
(' ', 9.106465816068027), (' ', 7.106465816068027), (' ',
5.106465816068027), (' ', 4.106465816068027), (' ', 2.1064658160680274),
(' ', 0.10646581606802741), (' ', 0.10646581606802741), (' ',
-2.8935341839319726), (' ', -2.8935341839319726), (' ', -3.8935341839319726)]
```

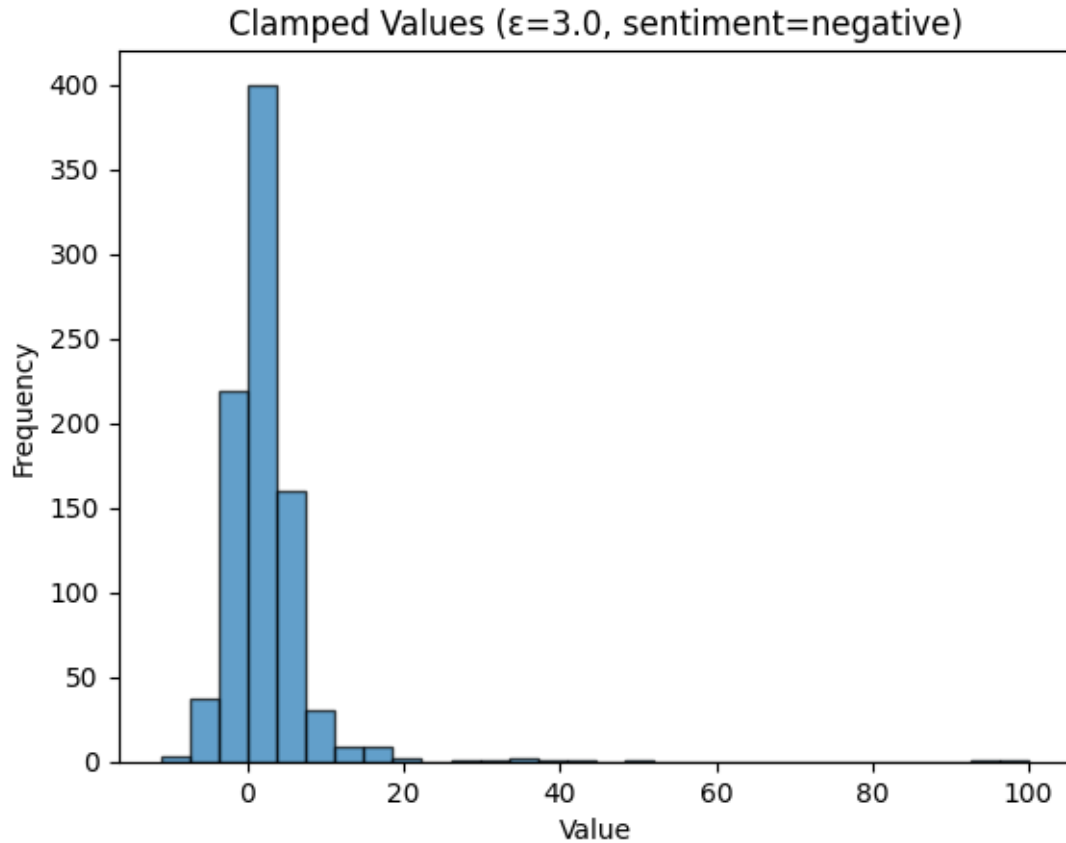




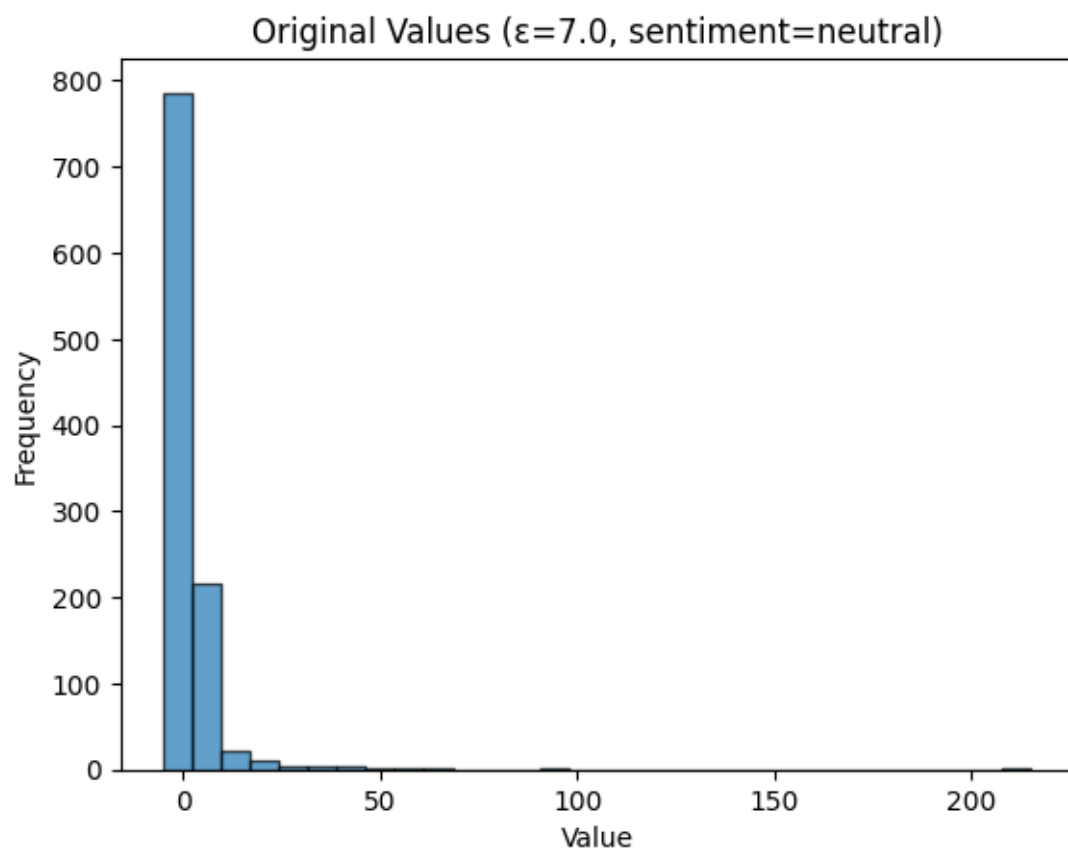
positive :

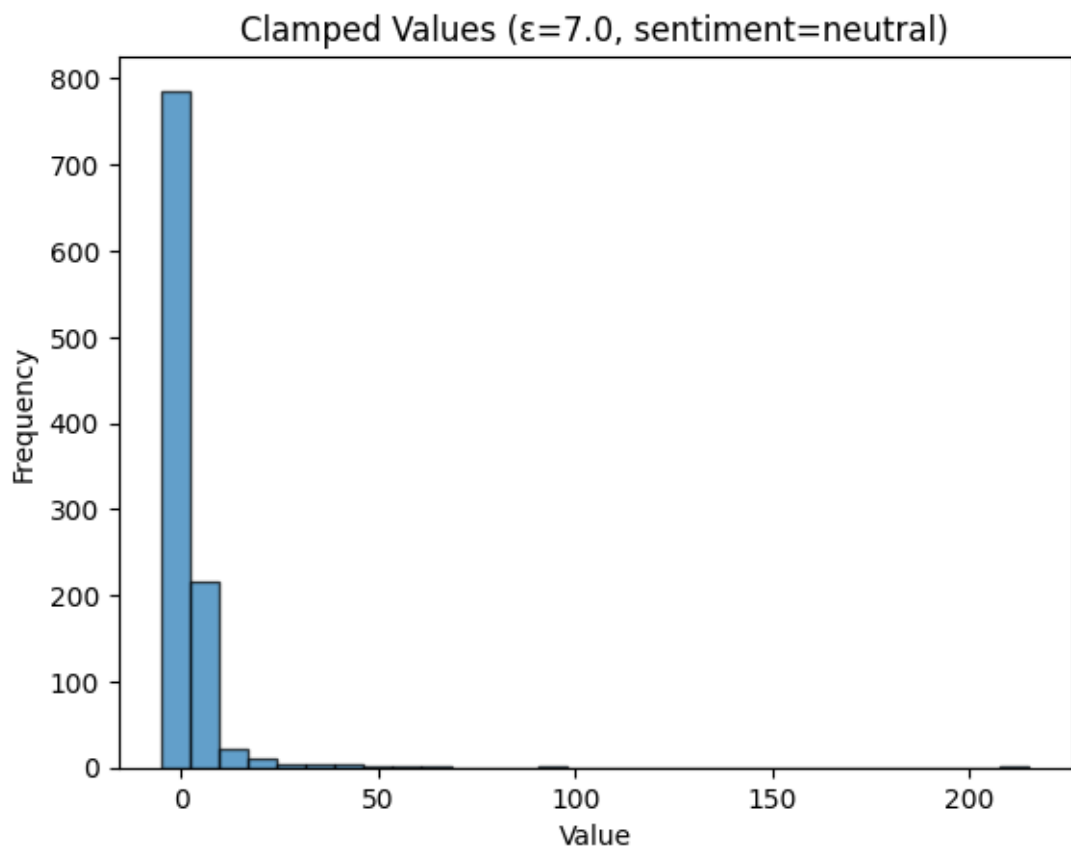
```
[(' ', 435.106465816068), (' ', 291.106465816068), (' ',
186.10646581606804), (' ', 175.10646581606804), (' ', 169.10646581606804),
(' ', 156.10646581606804), (' ', 152.10646581606804), (' ',
121.10646581606802), (' ', 118.10646581606802), (' ', 111.10646581606802),
(' ', 99.10646581606802), (' ', 78.10646581606802), (' ',
73.10646581606802), (' ', 73.10646581606802), (' ', 72.10646581606802), (' ',
65.10646581606802), (' ', 64.10646581606802), (' ', 64.10646581606802), (' ',
62.106465816068024), (' ', 57.106465816068024)]
```





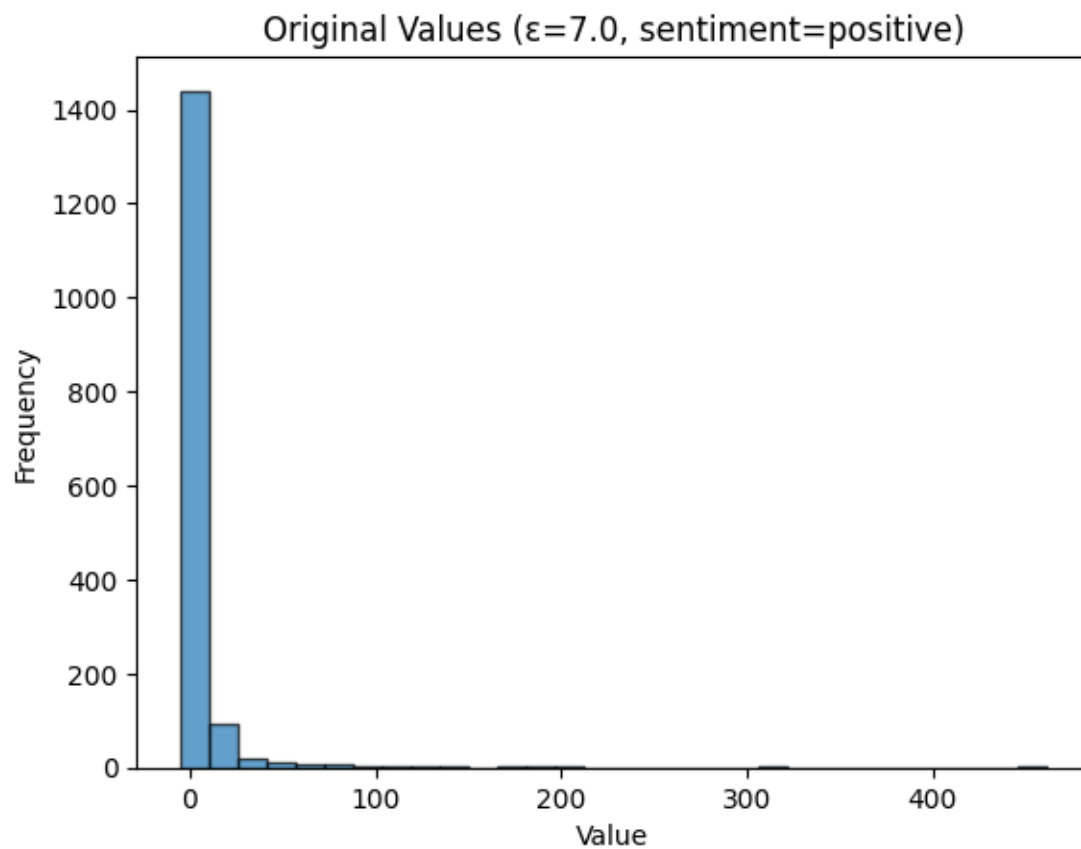
```
negative :
[(' ', 75.10646581606802), (' ', 71.10646581606802), (' ',
26.106465816068027), (' ', 16.106465816068027), (' ', 14.106465816068027),
(' ', 11.106465816068027), (' ', 10.106465816068027), (' ',
8.106465816068027), (' ', 2.1064658160680274), (' ', -2.8935341839319726),
(' ', -5.893534183931973), (' ', -6.893534183931973), (' ',
-6.893534183931973), (' ', -7.893534183931973), (' ', -7.893534183931973),
(' ', -8.893534183931973), (' ', -9.893534183931973), (' ',
-9.893534183931973), (' ', -9.893534183931973), (' ', -9.893534183931973)]
: 7.0
```

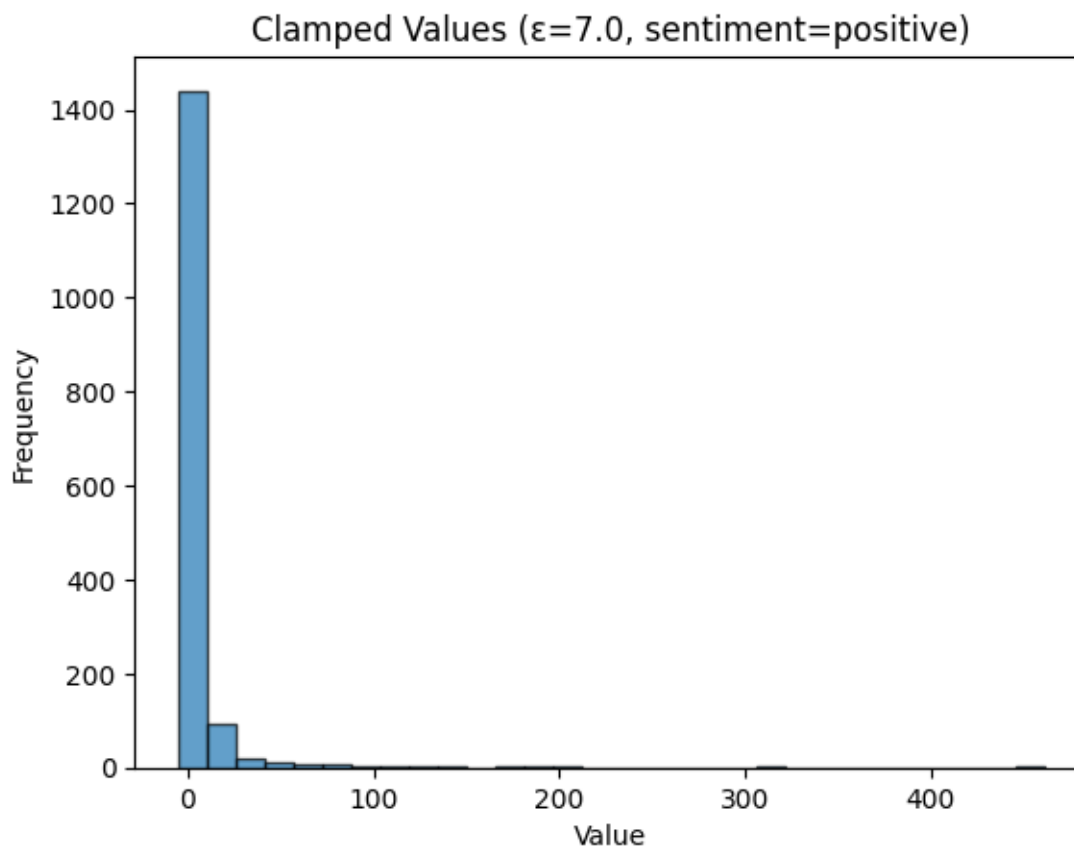




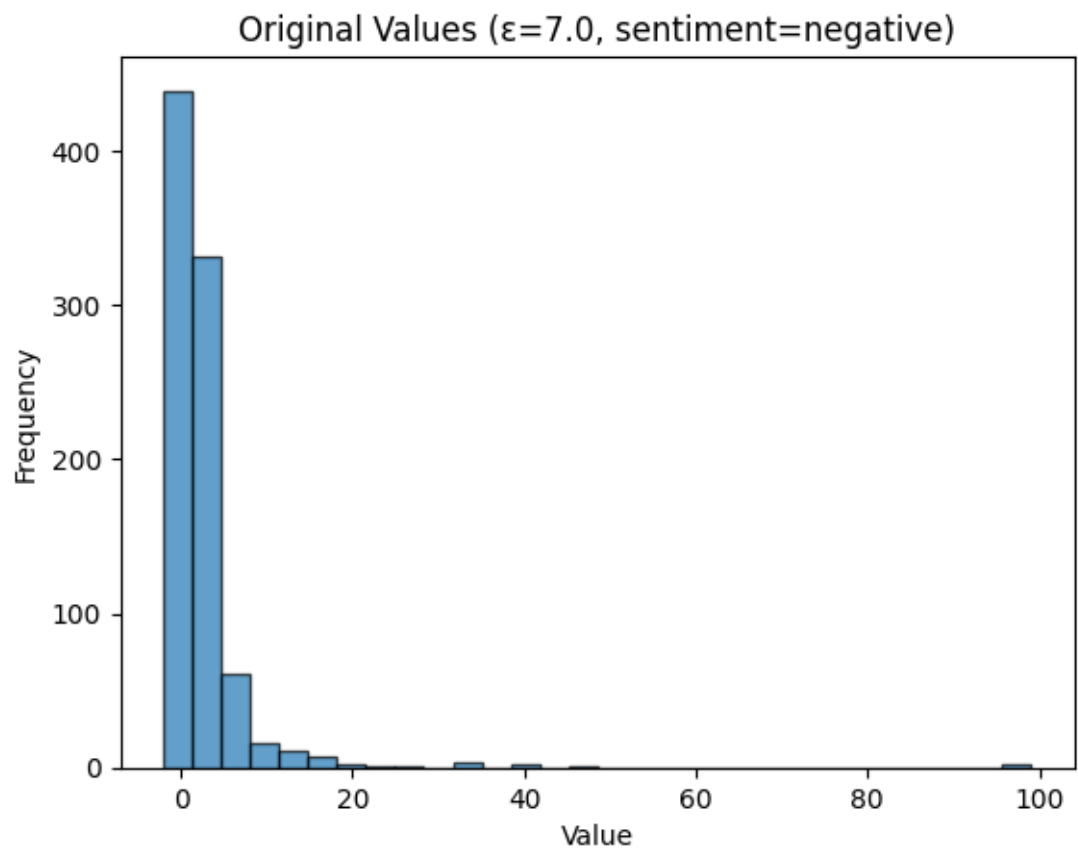
neutral :

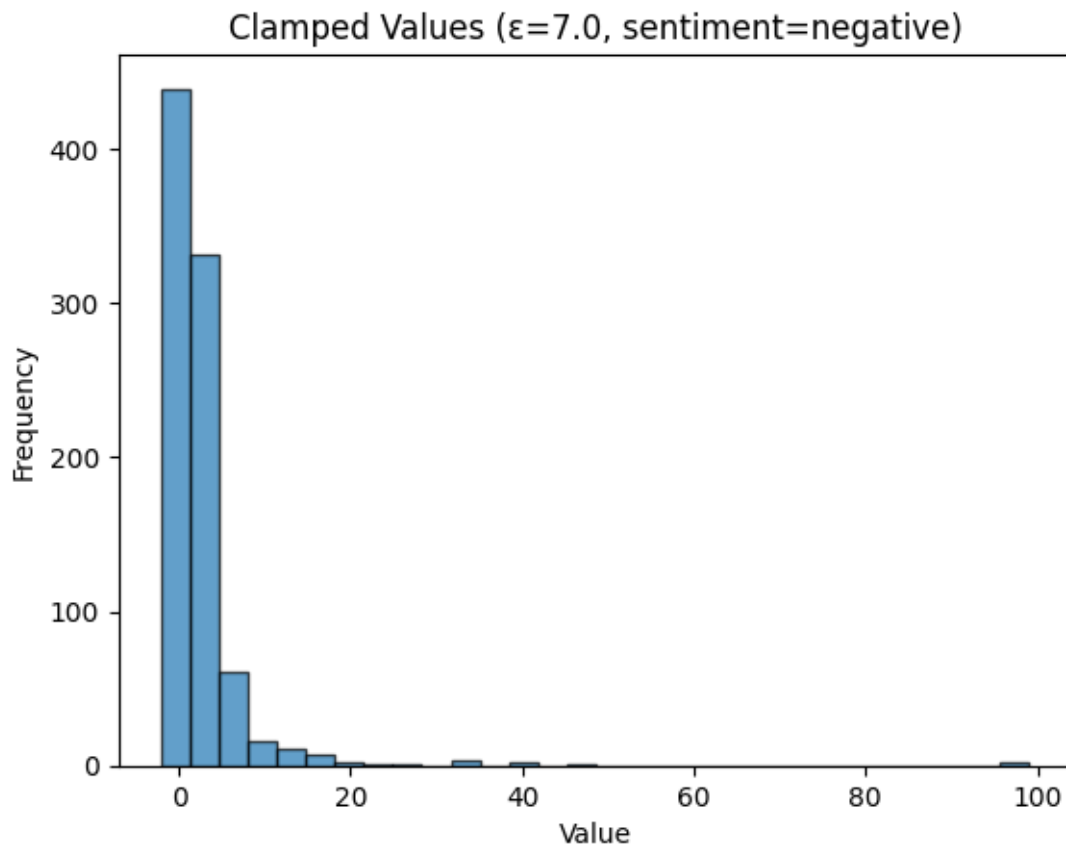
```
[(' ', 214.54405901722274), (' ', 93.54405901722274), (' ',
65.54405901722274), (' ', 60.54405901722274), (' ', 58.54405901722274), (' ',
49.54405901722274), (' ', 46.54405901722274), (' ', 45.54405901722274), (' ',
43.54405901722274), (' ', 42.54405901722274), (' ', 33.54405901722274),
(' ', 32.54405901722274), (' ', 32.54405901722274), (' ',
26.544059017222743), (' ', 25.544059017222743), (' ', 24.544059017222743),
(' ', 24.544059017222743), (' ', 21.544059017222743), (' ',
21.544059017222743), (' ', 21.544059017222743)]
```

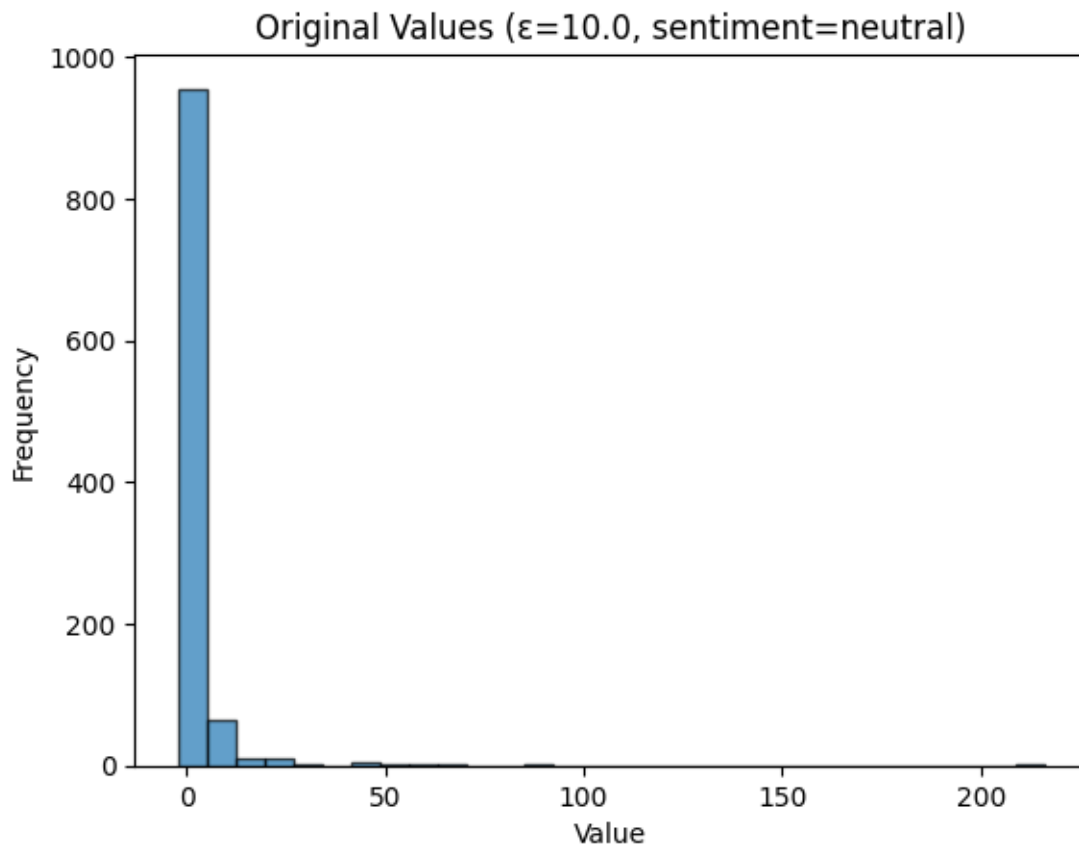


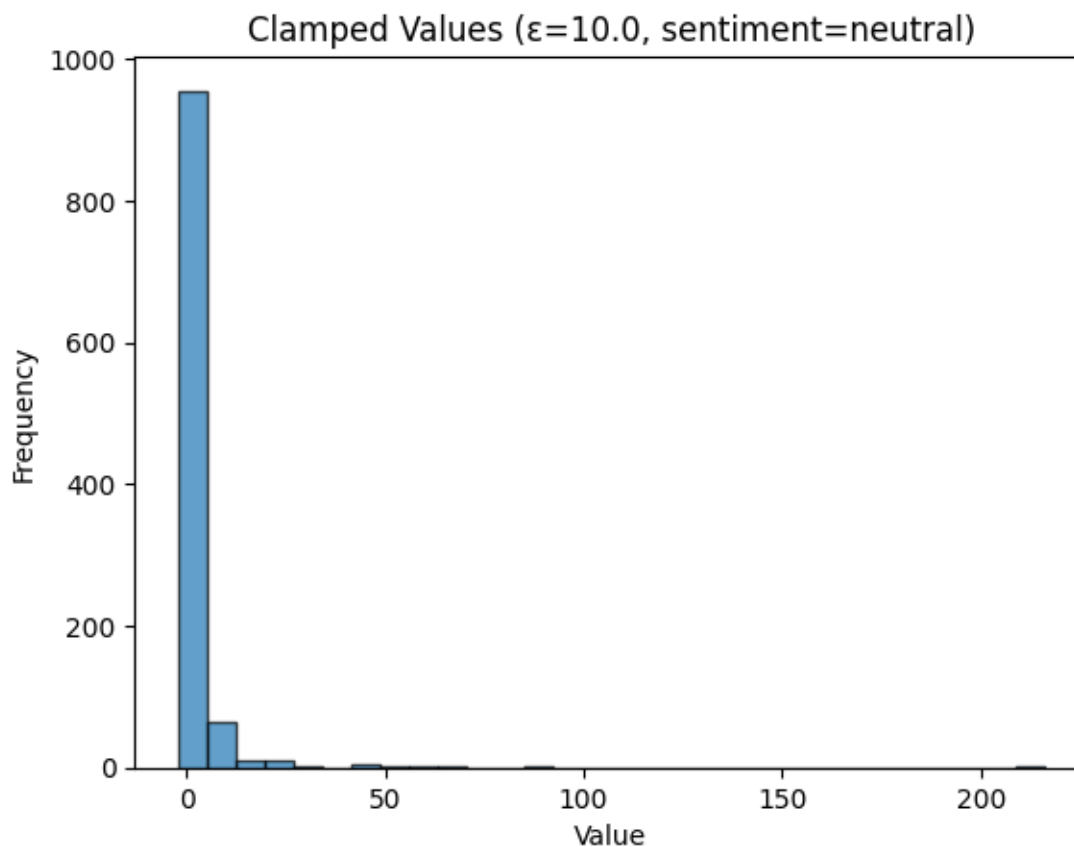
```
positive :
[(' ', 460.54405901722276), (' ', 315.54405901722276), (' ',
205.54405901722274), (' ', 199.54405901722274), (' ', 192.54405901722274),
(' ', 180.54405901722274), (' ', 175.54405901722274), (' ',
143.54405901722274), (' ', 140.54405901722274), (' ', 139.54405901722274),
(' ', 124.54405901722274), (' ', 105.54405901722274), (' ',
101.54405901722274), (' ', 97.54405901722274), (' ', 95.54405901722274),
(' ', 90.54405901722274), (' ', 90.54405901722274), (' ', 86.54405901722274),
(' ', 83.54405901722274), (' ', 79.54405901722274)]
```





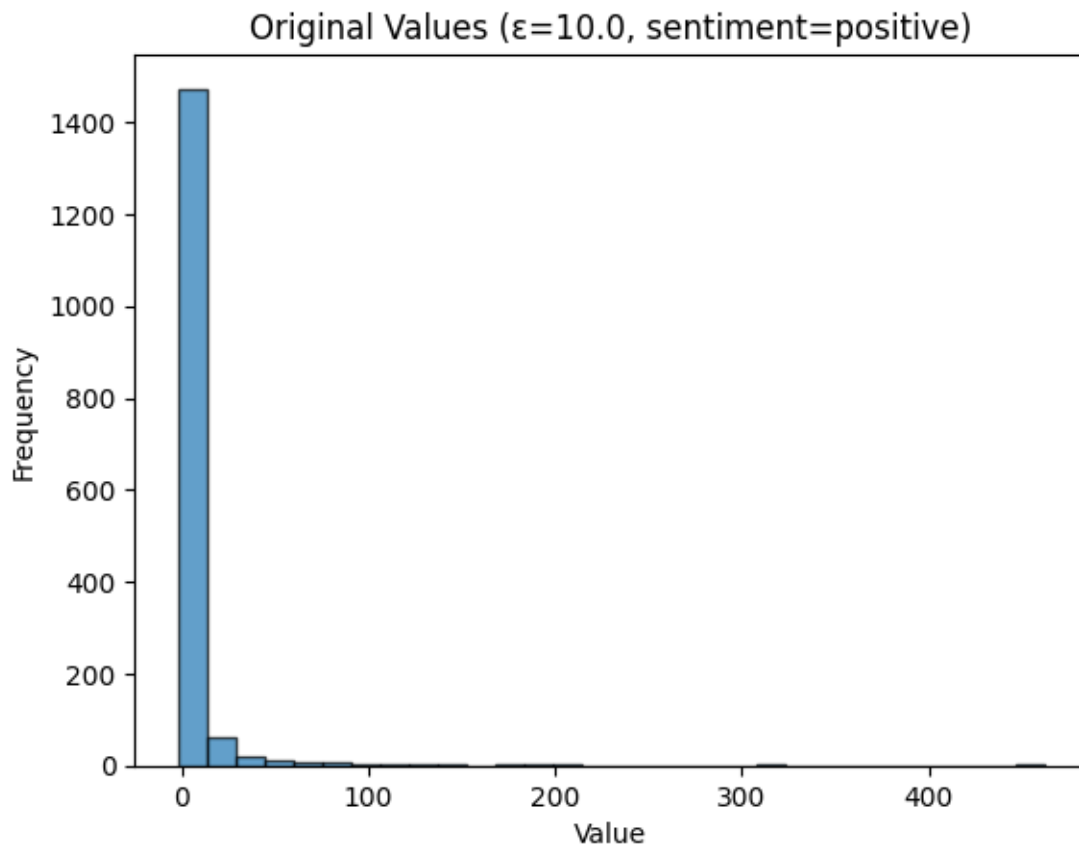
```
negative :
[(' ', 98.54405901722274), (' ', 98.54405901722274), (' ',
47.54405901722274), (' ', 40.54405901722274), (' ', 38.54405901722274), (' ',
33.54405901722274), (' ', 32.54405901722274), (' ', 31.544059017222743),
(' ', 25.544059017222743), (' ', 23.544059017222743), (' ',
20.544059017222743), (' ', 18.544059017222743), (' ', 16.544059017222743),
(' ', 15.544059017222741), (' ', 15.544059017222741), (' ',
15.544059017222741), (' ', 15.544059017222741), (' ', 14.544059017222741),
(' ', 14.544059017222741), (' ', 13.544059017222741)]
: 10.0
```

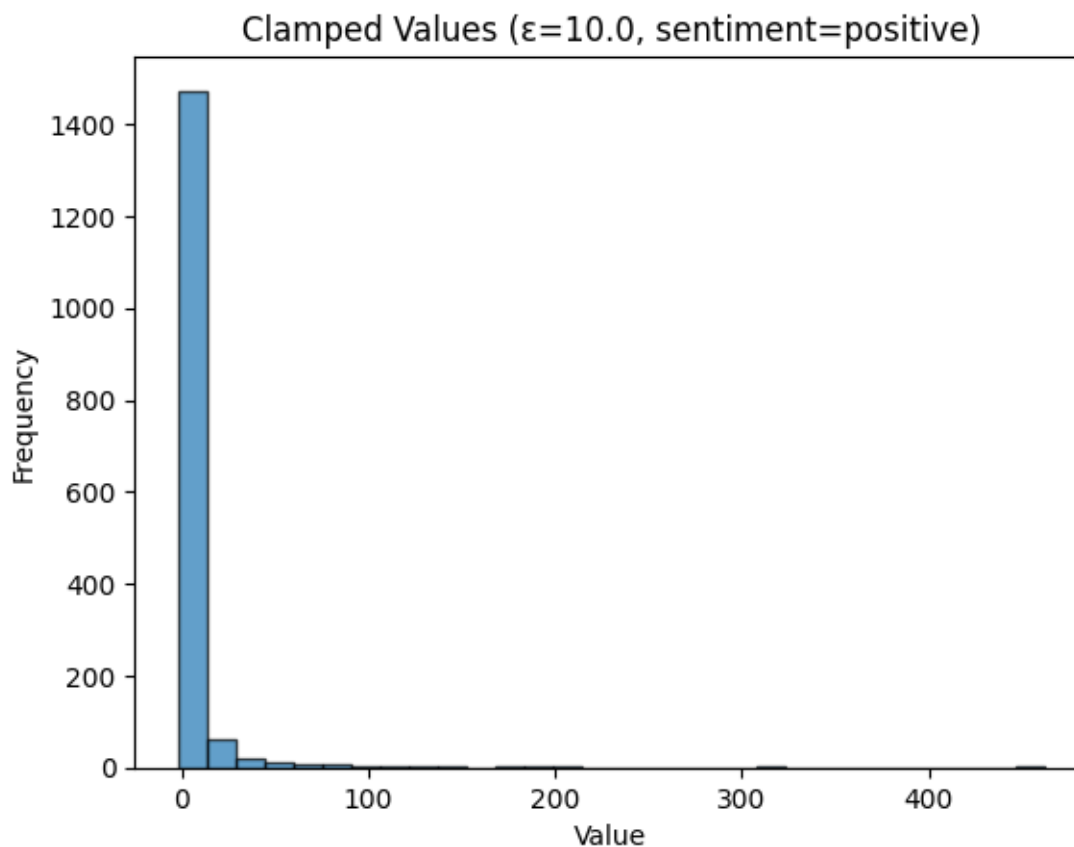




neutral :

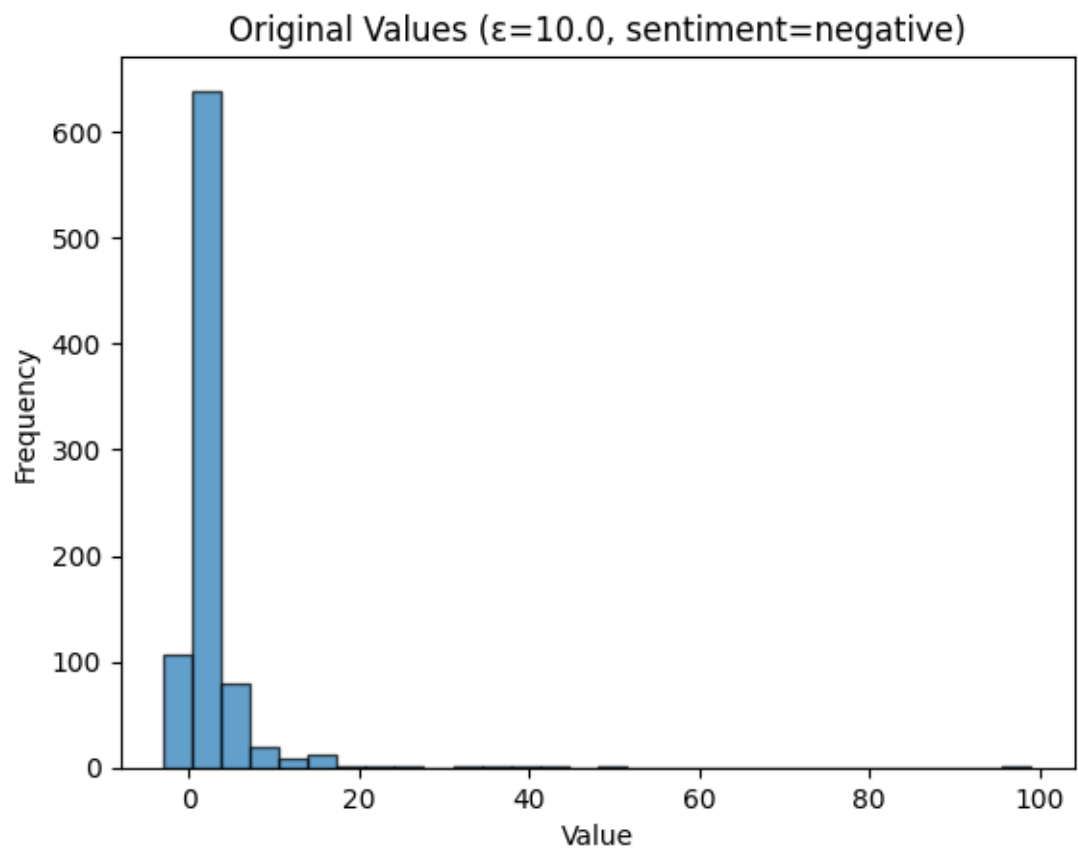
```
[(' ', 215.97730003511876), (' ', 91.97730003511876), (' ',
67.97730003511876), (' ', 59.97730003511876), (' ', 58.97730003511876), (' ',
49.97730003511876), (' ', 45.97730003511876), (' ', 44.97730003511876), (' ',
43.97730003511876), (' ', 42.97730003511876), (' ', 33.97730003511876),
(' ', 31.97730003511876), (' ', 30.97730003511876), (' ',
26.97730003511876), (' ', 25.97730003511876), (' ', 24.97730003511876), (' ',
24.97730003511876), (' ', 22.97730003511876), (' ', 21.97730003511876), (' ',
20.97730003511876)]
```

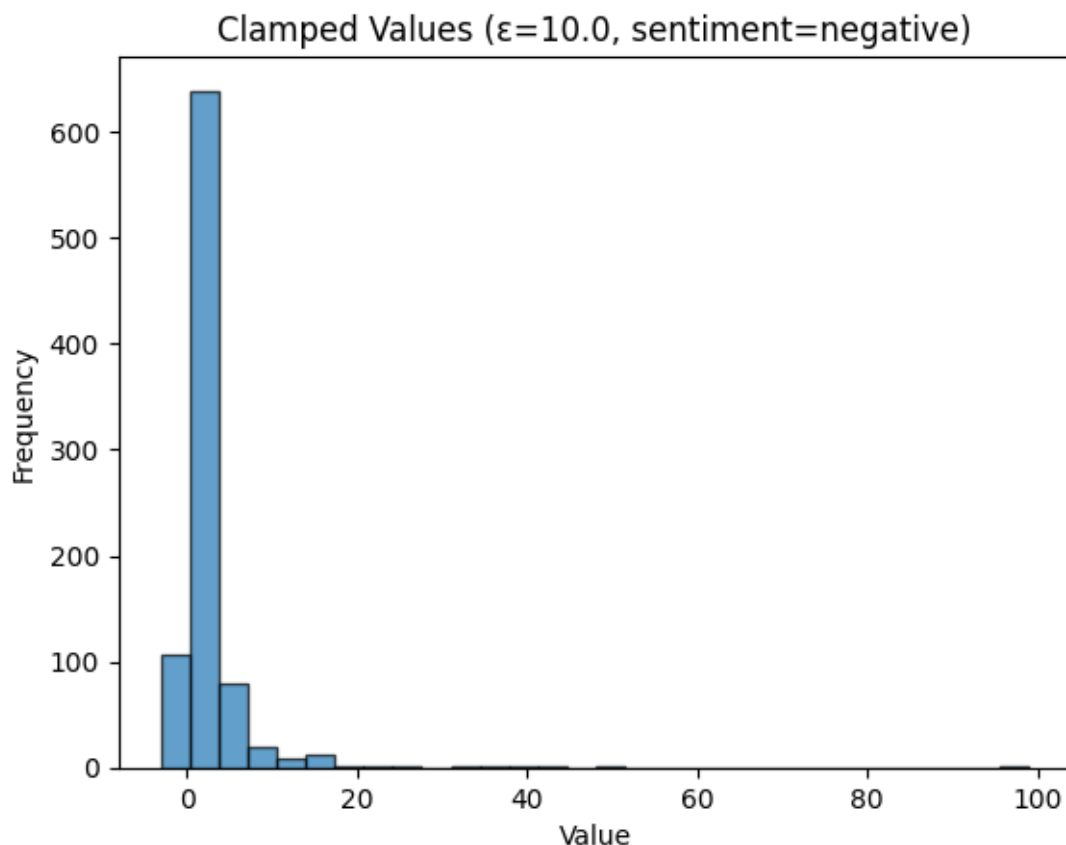




positive :

```
[(' ', 461.97730003511873), (' ', 314.97730003511873), (' ',
205.97730003511876), (' ', 197.97730003511876), (' ', 192.97730003511876),
(' ', 180.97730003511876), (' ', 174.97730003511876), (' ',
143.97730003511876), (' ', 140.97730003511876), (' ', 138.97730003511876),
(' ', 123.97730003511876), (' ', 106.97730003511876), (' ',
101.97730003511876), (' ', 98.97730003511876), (' ', 94.97730003511876),
(' ', 92.97730003511876), (' ', 88.97730003511876), (' ', 86.97730003511876),
(' ', 84.97730003511876), (' ', 79.97730003511876)]
```



negative :

```
[(' ', 98.97730003511876), (' ', 96.97730003511876), (' ',
47.97730003511876), (' ', 41.97730003511876), (' ', 38.97730003511876), (' ',
35.97730003511876), (' ', 33.97730003511876), (' ', 32.97730003511876), (' ',
25.97730003511876), (' ', 22.97730003511876), (' ', 21.97730003511876), (' ',
18.97730003511876), (' ', 15.977300035118757), (' ', 15.977300035118757),
(' ', 15.977300035118757), (' ', 15.977300035118757), (' ',
15.977300035118757), (' ', 14.977300035118757), (' ', 14.977300035118757),
(' ', 13.977300035118757)]
```

0.6 6.

```
[67]: #
for sentiment in df["sentiment"].unique():
    words_and_counts = top_k_words_and_counts(TOP_K,
term_frequencies[sentiment], vocabs[sentiment])
    results[sentiment]["no DP"] = words_and_counts

#
```

```

def display_results(results):
    for sentiment, data in results.items():
        print(f"\nSentiment: {sentiment}\n")
        df = pd.DataFrame(columns=["rank", "word", "count"] + [f"={eps}" for
↪eps in epsilons])
        for rank, (word, count) in enumerate(data["no DP"], start=1):
            row = {"rank": rank, "word": word, "count": count}
            for eps in epsilons:
                if eps in data and rank <= len(data[eps]):
                    row[f"={eps}"] = data[eps][rank-1][1] if rank-1 <
↪len(data[eps]) else None
                df = pd.concat([df, pd.DataFrame([row])], ignore_index=True)
            print(df)

display_results(results)

```

Sentiment: neutral

	rank	word	count	=0.01	=0.05	=0.1	=0.3	=0.7 \
0	1		216	4.975083	24.385288	-52.418709	-194.40911	-50.292652
1	2		93	4.975083	24.385288	-170.418709	-269.40911	-141.292652
2	3		67	4.975083	24.385288	-173.418709	-275.40911	-178.292652
3	4		60	4.975083	-32.614712	-178.418709	-280.40911	-179.292652
4	5		59	4.975083	-42.614712	-222.418709	-280.40911	-187.292652
5	6		50	4.975083	-57.614712	-223.418709	-283.40911	-190.292652
6	7		47	4.975083	-59.614712	-227.418709	-288.40911	-195.292652
7	8		45	4.975083	-72.614712	-229.418709	-289.40911	-198.292652
8	9		43	4.975083	-91.614712	-238.418709	-290.40911	-203.292652
9	10		43	4.975083	-92.614712	-243.418709	-290.40911	-205.292652
10	11		34	4.975083	-99.614712	-256.418709	-291.40911	-208.292652
11	12		32	4.975083	-108.614712	-258.418709	-298.40911	-209.292652
12	13		30	4.975083	-119.614712	-267.418709	-298.40911	-211.292652
13	14		27	4.975083	-119.614712	-277.418709	-299.40911	-213.292652
14	15		27	4.975083	-123.614712	-277.418709	-300.40911	-213.292652
15	16		25	4.975083	-126.614712	-279.418709	-301.40911	-215.292652
16	17		25	4.975083	-135.614712	-284.418709	-301.40911	-216.292652
17	18		23	4.975083	-135.614712	-284.418709	-302.40911	-217.292652
18	19		22	4.975083	-140.614712	-290.418709	-302.40911	-217.292652
19	20		21	4.975083	-149.614712	-293.418709	-303.40911	-217.292652

	=1.0	=2.0	=3.0	=7.0	=10.0
0	34.060279	148.332358	188.106466	214.544059	215.9773
1	-86.939721	23.332358	69.106466	93.544059	91.9773
2	-116.939721	0.332358	42.106466	65.544059	67.9773
3	-126.939721	-9.667642	35.106466	60.544059	59.9773
4	-126.939721	-12.667642	33.106466	58.544059	58.9773

5	-132.939721	-17.667642	26.106466	49.544059	49.9773
6	-134.939721	-19.667642	26.106466	46.544059	45.9773
7	-135.939721	-23.667642	20.106466	45.544059	44.9773
8	-138.939721	-24.667642	18.106466	43.544059	43.9773
9	-140.939721	-31.667642	18.106466	42.544059	42.9773
10	-141.939721	-31.667642	9.106466	33.544059	33.9773
11	-146.939721	-34.667642	7.106466	32.544059	31.9773
12	-151.939721	-37.667642	5.106466	32.544059	30.9773
13	-151.939721	-38.667642	4.106466	26.544059	26.9773
14	-151.939721	-38.667642	2.106466	25.544059	25.9773
15	-153.939721	-39.667642	0.106466	24.544059	24.9773
16	-153.939721	-41.667642	0.106466	24.544059	24.9773
17	-154.939721	-42.667642	-2.893534	21.544059	22.9773
18	-155.939721	-43.667642	-2.893534	21.544059	21.9773
19	-156.939721	-43.667642	-3.893534	21.544059	20.9773

Sentiment: positive

	rank	word	count	=0.01	=0.05	=0.1	=0.3	=0.7 \
0	1		463	4.975083	24.385288	47.581291	114.59089	222.707348
1	2		315	4.975083	24.385288	-80.418709	-54.40911	63.707348
2	3		206	4.975083	24.385288	-121.418709	-160.40911	-42.292652
3	4		198	4.975083	14.385288	-132.418709	-172.40911	-54.292652
4	5		193	4.975083	14.385288	-149.418709	-190.40911	-64.292652
5	6		181	4.975083	-3.614712	-156.418709	-194.40911	-69.292652
6	7		175	4.975083	-42.614712	-173.418709	-207.40911	-75.292652
7	8		144	4.975083	-42.614712	-199.418709	-212.40911	-86.292652
8	9		141	4.975083	-49.614712	-203.418709	-231.40911	-98.292652
9	10		140	4.975083	-59.614712	-209.418709	-235.40911	-108.292652
10	11		124	4.975083	-66.614712	-214.418709	-238.40911	-122.292652
11	12		106	4.975083	-67.614712	-220.418709	-254.40911	-137.292652
12	13		102	4.975083	-73.614712	-220.418709	-266.40911	-142.292652
13	14		99	4.975083	-78.614712	-220.418709	-269.40911	-142.292652
14	15		96	4.975083	-84.614712	-221.418709	-275.40911	-149.292652
15	16		90	4.975083	-89.614712	-227.418709	-277.40911	-150.292652
16	17		90	4.975083	-98.614712	-235.418709	-279.40911	-151.292652
17	18		87	4.975083	-100.614712	-242.418709	-280.40911	-161.292652
18	19		84	4.975083	-100.614712	-247.418709	-280.40911	-163.292652
19	20		80	4.975083	-109.614712	-247.418709	-283.40911	-165.292652

	=1.0	=2.0	=3.0	=7.0	=10.0
0	279.060279	397.332358	435.106466	460.544059	461.9773
1	122.060279	250.332358	291.106466	315.544059	314.9773
2	32.060279	138.332358	186.106466	205.544059	205.9773
3	24.060279	130.332358	175.106466	199.544059	197.9773
4	9.060279	124.332358	169.106466	192.544059	192.9773
5	-2.939721	119.332358	156.106466	180.544059	180.9773
6	-10.939721	110.332358	152.106466	175.544059	174.9773

7	-41.939721	79.332358	121.106466	143.544059	143.9773
8	-43.939721	72.332358	118.106466	140.544059	140.9773
9	-49.939721	66.332358	111.106466	139.544059	138.9773
10	-54.939721	53.332358	99.106466	124.544059	123.9773
11	-75.939721	40.332358	78.106466	105.544059	106.9773
12	-78.939721	35.332358	73.106466	101.544059	101.9773
13	-84.939721	31.332358	73.106466	97.544059	98.9773
14	-87.939721	29.332358	72.106466	95.544059	94.9773
15	-92.939721	21.332358	65.106466	90.544059	92.9773
16	-92.939721	20.332358	64.106466	90.544059	88.9773
17	-95.939721	19.332358	64.106466	86.544059	86.9773
18	-97.939721	18.332358	62.106466	83.544059	84.9773
19	-102.939721	12.332358	57.106466	79.544059	79.9773

Sentiment: negative

	rank	word count	=0.01	=0.05	=0.1	=0.3	=0.7 \
0	1	99	4.975083	24.385288	-161.418709	-255.40911	-151.292652
1	2	97	4.975083	24.385288	-177.418709	-264.40911	-158.292652
2	3	48	4.975083	24.385288	-213.418709	-271.40911	-192.292652
3	4	42	4.975083	24.385288	-224.418709	-279.40911	-202.292652
4	5	39	4.975083	-20.614712	-239.418709	-293.40911	-203.292652
5	6	36	4.975083	-24.614712	-252.418709	-297.40911	-203.292652
6	7	34	4.975083	-29.614712	-258.418709	-300.40911	-204.292652
7	8	33	4.975083	-31.614712	-259.418709	-301.40911	-204.292652
8	9	27	4.975083	-45.614712	-260.418709	-308.40911	-205.292652
9	10	24	4.975083	-46.614712	-264.418709	-313.40911	-209.292652
10	11	22	4.975083	-93.614712	-266.418709	-315.40911	-211.292652
11	12	19	4.975083	-105.614712	-275.418709	-316.40911	-211.292652
12	13	16	4.975083	-107.614712	-282.418709	-316.40911	-212.292652
13	14	16	4.975083	-127.614712	-296.418709	-316.40911	-213.292652
14	15	16	4.975083	-131.614712	-297.418709	-316.40911	-213.292652
15	16	16	4.975083	-145.614712	-299.418709	-317.40911	-214.292652
16	17	15	4.975083	-146.614712	-307.418709	-318.40911	-214.292652
17	18	15	4.975083	-152.614712	-309.418709	-319.40911	-214.292652
18	19	15	4.975083	-152.614712	-311.418709	-319.40911	-216.292652
19	20	14	4.975083	-169.614712	-313.418709	-320.40911	-217.292652

		=1.0	=2.0	=3.0	=7.0	=10.0
0	-81.939721	31.332358	75.106466	98.544059	98.9773	
1	-82.939721	29.332358	71.106466	98.544059	96.9773	
2	-133.939721	-4.667642	26.106466	47.544059	47.9773	
3	-138.939721	-25.667642	16.106466	40.544059	41.9773	
4	-139.939721	-25.667642	14.106466	38.544059	38.9773	
5	-144.939721	-26.667642	11.106466	33.544059	35.9773	
6	-145.939721	-32.667642	10.106466	32.544059	33.9773	
7	-146.939721	-36.667642	8.106466	31.544059	32.9773	
8	-149.939721	-38.667642	2.106466	25.544059	25.9773	

```

9 -151.939721 -42.667642 -2.893534 23.544059 22.9773
10 -151.939721 -42.667642 -5.893534 20.544059 21.9773
11 -151.939721 -44.667642 -6.893534 18.544059 18.9773
12 -152.939721 -45.667642 -6.893534 16.544059 15.9773
13 -153.939721 -46.667642 -7.893534 15.544059 15.9773
14 -154.939721 -47.667642 -7.893534 15.544059 15.9773
15 -155.939721 -48.667642 -8.893534 15.544059 15.9773
16 -155.939721 -48.667642 -9.893534 15.544059 15.9773
17 -157.939721 -49.667642 -9.893534 14.544059 14.9773
18 -157.939721 -50.667642 -9.893534 14.544059 14.9773
19 -158.939721 -50.667642 -9.893534 13.544059 13.9773

```

/tmp/ipykernel_931337/446711119.py:17: FutureWarning: The behavior of DataFrame concatenation with empty or all-NA entries is deprecated. In a future version, this will no longer exclude empty or all-NA columns when determining the result dtypes. To retain the old behavior, exclude the relevant entries before the concat operation.

```
df = pd.concat([df, pd.DataFrame([row])], ignore_index=True)
```

/tmp/ipykernel_931337/446711119.py:17: FutureWarning: The behavior of DataFrame concatenation with empty or all-NA entries is deprecated. In a future version, this will no longer exclude empty or all-NA columns when determining the result dtypes. To retain the old behavior, exclude the relevant entries before the concat operation.

```
df = pd.concat([df, pd.DataFrame([row])], ignore_index=True)
```

/tmp/ipykernel_931337/446711119.py:17: FutureWarning: The behavior of DataFrame concatenation with empty or all-NA entries is deprecated. In a future version, this will no longer exclude empty or all-NA columns when determining the result dtypes. To retain the old behavior, exclude the relevant entries before the concat operation.

```
df = pd.concat([df, pd.DataFrame([row])], ignore_index=True)
```

0.7 7.

```

[68]: #
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm

#
font_path = '/usr/share/fonts/truetype/fonts-japanese-gothic.ttf' #
font_prop = fm.FontProperties(fname=font_path)

def calculate_match_rate(original_top_k, dp_top_k):
    match_count = len(set(original_top_k) & set(dp_top_k))
    return match_count / len(original_top_k) if original_top_k else 0

def plot_match_rate(results, epsilons, sentiment_label):
    top_k_values = [3, 5, 10, 20]
    match_rates = {k: [] for k in top_k_values}

```

```

for eps in epsilons:
    for k in top_k_values:
        original_top_k = [word for word, count in
↪results[sentiment_label]["no DP"][:k]]
        dp_top_k = [word for word, count in results[sentiment_label][eps][:
↪k]]

        match_rate = calculate_match_rate(original_top_k, dp_top_k)
        match_rates[k].append(match_rate)

plt.figure(figsize=(10, 6))
for k, rates in match_rates.items():
    plt.plot(epsilons, rates, label=f'top{k}')

plt.xlabel(' ', fontproperties=font_prop)
plt.ylabel(' ', fontproperties=font_prop)
plt.title(f'({sentiment_label})', fontproperties=font_prop)
plt.legend(prop=font_prop)
plt.grid(True)
plt.show()

epsilons = [0.01, 0.05, 0.1, 0.3, 0.7, 1.0, 2.0, 3.0, 7.0, 10.0]

plot_match_rate(results, epsilons, 'negative')
plot_match_rate(results, epsilons, 'neutral')
plot_match_rate(results, epsilons, 'positive')

```

