NLP -02, I have a 66 MB dataset of movie reviews, and I'll be using text processing techniques like tokenization, stopword removal, lemmatization, and sentiment analysisto extract insights and analyze the overall sentiment (positive, negative, neutral) of the reviews.

```
import nltk
import pandas as pd
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from nltk.sentiment import SentimentIntensityAnalyzer

# Download necessary NLTK resources
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('vader_lexicon')
nltk.download('punkt_tab')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Package punkt_tab is already up-to-date!
```

```
True
```

```
df = pd.read_csv("/content/drive/MyDrive/NLP/IMDB Dataset.csv")
```

```
print(df.head())
```

```
                                              review sentiment
0  One of the other reviewers has mentioned that ...  positive
1  A wonderful little production. <br /><br />The...  positive
2  I thought this was a wonderful way to spend ti...  positive
3  Basically there's a family where a little boy ...  negative
4  Petter Mattei's "Love in the Time of Money" is...  positive
```

```
df['tokens'] = df['review'].apply(lambda x: word_tokenize(x.lower()))

stop_words = set(stopwords.words('english'))
df['filtered_tokens'] = df['tokens'].apply(lambda x: [word for word in
x if word.isalnum() and word not in stop_words])

lemmatizer = WordNetLemmatizer()
df['lemmatized_tokens'] = df['filtered_tokens'].apply(lambda x:
[lemmatizer.lemmatize(word) for word in x])
```

```
sia = SentimentIntensityAnalyzer()
df['sentiment_score'] = df['review'].apply(lambda x:
sia.polarity_scores(x)['compound'])

print(df[['review', 'tokens', 'filtered_tokens', 'lemmatized_tokens',
'sentiment_score']])
```

```
                                                  review  \
0      One of the other reviewers has mentioned that ...
1      A wonderful little production. <br /><br />The...
2      I thought this was a wonderful way to spend ti...
3      Basically there's a family where a little boy ...
4      Petter Mattei's "Love in the Time of Money" is...
...                                                  ...
49995  I thought this movie did a down right good job...
49996  Bad plot, bad dialogue, bad acting, idiotic di...
49997  I am a Catholic taught in parochial elementary...
49998  I'm going to have to disagree with the previou...
49999  No one expects the Star Trek movies to be high...

                                                  tokens  \
0      [one, of, the, other, reviewers, has, mentione...
1      [a, wonderful, little, production, ., <, br, /...
2      [i, thought, this, was, a, wonderful, way, to,...
3      [basically, there, 's, a, family, where, a, li...
4      [petter, mattei, 's, ``, love, in, the, time, ...
...                                                  ...
49995  [i, thought, this, movie, did, a, down, right,...
49996  [bad, plot, ,, bad, dialogue, ,, bad, acting, ...
49997  [i, am, a, catholic, taught, in, parochial, el...
49998  [i, 'm, going, to, have, to, disagree, with, t...
49999  [no, one, expects, the, star, trek, movies, to...

                                          filtered_tokens  \
0      [one, reviewers, mentioned, watching, 1, oz, e...
1      [wonderful, little, production, br, br, filmin...
2      [thought, wonderful, way, spend, time, hot, su...
3      [basically, family, little, boy, jake, thinks,...
4      [petter, mattei, love, time, money, visually, ...
...                                                  ...
49995  [thought, movie, right, good, job, creative, o...
49996  [bad, plot, bad, dialogue, bad, acting, idioti...
49997  [catholic, taught, parochial, elementary, scho...
49998  [going, disagree, previous, comment, side, mal...
49999  [one, expects, star, trek, movies, high, art, ...

                                       lemmatized_tokens
sentiment_score
0      [one, reviewer, mentioned, watching, 1, oz, ep...          -
0.9951
```

```
1      [wonderful, little, production, br, br, filmin...
0.9641
2      [thought, wonderful, way, spend, time, hot, su...
0.9605
3      [basically, family, little, boy, jake, think, ...           -
0.9213
4      [petter, mattei, love, time, money, visually, ...
0.9744
...                                                  ...
...
49995  [thought, movie, right, good, job, creative, o...
0.9890
49996  [bad, plot, bad, dialogue, bad, acting, idioti...           -
0.6693
49997  [catholic, taught, parochial, elementary, scho...           -
0.9851
49998  [going, disagree, previous, comment, side, mal...           -
0.7648
49999  [one, expects, star, trek, movie, high, art, f...
0.4329

[50000 rows x 5 columns]
```