488-NLP Sentiment Analysis

```python
In [1]:  import pandas as pd
         import re
         import nltk
         from nltk.corpus import stopwords
         from nltk.stem import WordNetLemmatizer
```

```python
In [2]:  nltk.download('stopwords')
         nltk.download('wordnet')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data]   Unzipping corpora\stopwords.zip.
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
```

```
Out[2]:  True
```

```python
In [5]:  file_path = r"file:///C:/Users/HP/Downloads/Restaurant_Reviews.tsv"
         data = pd.read_csv(file_path, delimiter='\t')
```

```python
In [6]:  lemmatizer = WordNetLemmatizer()
         stop_words = set(stopwords.words('english'))
```

```python
In [7]:  def preprocess_text(text):
             # Remove special characters, numbers, and punctuation
             text = re.sub(r'[^a-zA-Z\s]', '', text)
             # Convert to lowercase
             text = text.lower()
             # Tokenize the text
             tokens = text.split()
             # Remove stopwords and lemmatize each token
             lemmatized_tokens = [lemmatizer.lemmatize(token) for token in tokens if token r
             return ' '.join(lemmatized_tokens), lemmatized_tokens
```

```python
In [8]:  import nltk
         nltk.download('omw-1.4')
         nltk.download('wordnet')  # Ensures WordNet itself is also available.
```

```
[nltk_data] Downloading package omw-1.4 to
[nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\HP\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
Out[8]:  True
```

```python
In [9]:  from nltk.corpus import stopwords
         from nltk.stem import WordNetLemmatizer

         # Ensure the required corpora are downloaded
         nltk.download('stopwords')
         nltk.download('wordnet')

         # Initialize resources
         stop_words = set(stopwords.words('english'))
         lemmatizer = WordNetLemmatizer()
```

```python
def preprocess_text(text):
    tokens = text.split()
    lemmatized_tokens = [lemmatizer.lemmatize(token) for token in tokens if token.]
    return ' '.join(lemmatized_tokens), lemmatized_tokens
```

In [10]: `data['Cleaned_Review'], data['Lemmatized_Tokens'] = zip(*data['Review'].apply(prepr`

In [11]: `print(data[['Review', 'Cleaned_Review', 'Lemmatized_Tokens']].head())`

```
                                              Review  \
0                              Wow... Loved this place.
1                                    Crust is not good.
2             Not tasty and the texture was just nasty.
3   Stopped by during the late May bank holiday of...
4   The selection on the menu was great and so wer...

                                      Cleaned_Review  \
0                                  Wow... Loved place.
1                                          Crust good.
2                                  tasty texture nasty.
3   Stopped late May bank holiday Rick Steve recom...
4                          selection menu great prices.

                                    Lemmatized_Tokens
0                              [Wow..., Loved, place.]
1                                        [Crust, good.]
2                             [tasty, texture, nasty.]
3   [Stopped, late, May, bank, holiday, Rick, Stev...
4                  [selection, menu, great, prices.]
```

In [12]: `print(data[['Review', 'Cleaned_Review', 'Lemmatized_Tokens']].tail())`

```
                                              Review  \
995  I think food should have flavor and texture an...
996                              Appetite instantly gone.
997  Overall I was not impressed and would not go b...
998  The whole experience was underwhelming, and I ...
999  Then, as if I hadn't wasted enough of my life ...

                                      Cleaned_Review  \
995                      think food flavor texture lacking.
996                              Appetite instantly gone.
997                      Overall impressed would go back.
998  whole experience underwhelming, think we'll go...
999  Then, wasted enough life there, poured salt wo...

                                    Lemmatized_Tokens
995          [think, food, flavor, texture, lacking.]
996                      [Appetite, instantly, gone.]
997            [Overall, impressed, would, go, back.]
998  [whole, experience, underwhelming,, think, we'...
999  [Then,, wasted, enough, life, there,, poured, ...
```

In [ ]: