

```
import numpy as np
import pandas as pd
import nltk
import re
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer

# 1. Load the dataset
# Make sure your Drive is mounted!
data = pd.read_csv('/content/drive/MyDrive/Restaurant_Reviews.tsv', delimiter='\t', quoting=3)

# 2. Download stopwords
nltk.download('stopwords')
all_stopwords = stopwords.words('english')
# Optional: 'not' is a stopword but important for sentiment.
# You might want to remove it from the stopword list:
# all_stopwords.remove('not')

# 3. Pre-processing (The Corpus)
corpus = []
ps = PorterStemmer()

for i in range(0, 1000):
    # Remove non-letters
    review = re.sub('[^a-zA-Z]', ' ', data['Review'][i])
    # Lowercase and split
    review = review.lower().split()
    # Stemming and removing stopwords
    review = [ps.stem(word) for word in review if not word in set(all_stopwords)]
    # Rejoin into a string
    review = ' '.join(review)
    corpus.append(review)

# View the first 5 processed results
print(corpus[:5])
```