1. Choosing the data:

a) Choose existing large documents from NLTK or from the Gutenberg collection on the web, or b) Collect your own data, by using your own documents or collecting data from other sources. Combine the text from these sources to make two documents for the corpora for the first task. Describe the method that you used to define and collect the data, including the difference between the documents. Note any limitations to the method or the text that you were able to find. Do preprocessing to get the text in a suitable format for processing and describe what you did.

Ans) I selected data from a dataset containing Google reviews for restaurants in the USA. The dataset is in JSON format and is quite large. Parsing the data to extract the review data for two specific restaurants posed a significant challenge. My approach involved using the unique business_id to identify and merge the data related to the two restaurants. The goal was to analyze the top two restaurants with the highest number of reviews.

To understand the structure and fields of the JSON object, I developed a JSON parser. This parser allowed me to examine the schema of the JSON file, which was crucial for further processing. With the schema in hand, I proceeded to identify the top two most reviewed restaurants based on their unique business_id.

Finally, I created two arrays to store the reviews associated with each restaurant. These arrays served as the foundation for my subsequent analysis.

```
In [90]:  #preprocessing data and collecting two files
          import requests
          import json

          response = requests.get('https://datarepo.eng.ucsd.edu/mcauley_group/gdrive/
          response_json= json.loads(response.text)
          # print(response_json["train"][0])
          dic= {};

          # Schema buildig

          for x in response_json["train"]:
              for k,v in x.items():

                  if(k=="business_id"):
```

```python
            if(v in dic):
                a =dic[v]
                dic[v]=a+1
            else:
                dic[v]=1




total_no_reviews=[]
#highest reviewed places
for k,v in dic.items():
    total_no_reviews.append(v)

total_no_reviews.sort(reverse=True)




#top_two would store the business_id of the two restaurants
top_two=[]

for k,v in dic.items():
    if(v==total_no_reviews[0] or v==total_no_reviews[1]):
        top_two.append(k)

#the below two lists will store all the reviews of first and second restaura
reviews_one=[]
reviews_two=[]
print("Business_id of top two restaurants")
print(top_two)
#iterating the json object to retrieve all the reviews
f1= False
f2 =False

for x in response_json["train"]:
    for k,v in x.items():

        if(k=="business_id" and v==top_two[0]):
            f1=True
        elif (k=="business_id" and v==top_two[1]):
            f2=True
        elif(k=="business_id"):
            f1=False
            f2=False
            break
        if(f1==True and k=="review_text"):
            f1=False
            reviews_one.append(v)
        if(f2==True and k=="review_text"):
            f2=False
```

```python
            reviews_two.append(v)

first_review = []
second_review = []

# tokenization : processing the data the way we want i.e separated with all
for i in range(len(reviews_one)):
    x = reviews_one[i]
    y = x.split(" ")
    for j in range(len(y)):
        first_review.append(y[j])

for i in range(len(reviews_two)):
    x = reviews_two[i]
    y = x.split(" ")
    for j in range(len(y)):
        second_review.append(y[j])
```

```
Business_id of top two restaurants
['6043ad17b81264dfa846c9ea', '604245d6b9a6829e686e8c2a']
```

Preprocessing and Reasons for Preprocessing

I chose the following processing options for the text data:

1. Lowercase Conversion: Converting all words to lowercase helped me to standardize the text and treat words with different cases as the same word. It reduced the vocabulary size and ensured that words with the same meaning but different cases are treated as identical, improving the accuracy of subsequent analysis.

2. Stopwords Removal: The ccommonly occurring stop words (e.g., "a," "the," "is") that do not carry significant meaning and are often removed to focus on more informative words. By removing stopwords, I eliminated noise and reduced the dimensionality of the data, allowing me to focus on more relevant and meaningful words. I also added a few other stop words to remove the other words not provided by the nltk package.

3. Lemmatization: Lemmatization reduces words to their base or root form, such as converting "running" to "run" or "cats" to "cat." It helps to normalize the words and reduces the variation in word forms, enabling us to capture the essential meaning of words while reducing sparsity in the data. This can improve the accuracy of downstream tasks like text classification or topic modeling. So I just wanted to see the word in the most basic form as possible.

To help the eliminate noise, standardize text, and focus on the most informative aspects of the data.

```
In [71]: import nltk
         import re
         # this regular expression pattern matches any word that contains all non-alp
         #    lower-case characters [^a-z]+
         # the beginning ^ and ending $ require the match to begin and end on a word
         # def alpha_filter(w):
         #     # pattern to match a word of non-alphabetical characters
         #     pattern = r"\b(?![!\"#$%&()*+-./:;<=>?@[\]^_`{|}~\n])\w+\b"

         # # Find all matching words
         # #     matches = re.findall(pattern, w)
         #     return bool(re.match(pattern, w))
         def alpha_filter(w):
           pattern = re.compile('^[^a-z]+$')
           if (pattern.match(w)):
             return True
           else:
             return False

         alpha_reviews_one = [x for x in first_review if not alpha_filter(x)]
         alpha_reviews_two = [x for x in second_review if  not alpha_filter(x)]
         print(alpha_reviews_one)
```

```
['Awesome', 'pizza', '(Original).', 'Hammer', 'Pizza', 'und', 'sie', 'schmeck
t', 'einfach', 'kstlich.', 'Hammer', 'Pizza', 'and', 'it', 'just', 'tastes',
'delicious.', 'Also', 'wenn', 'man', 'in', 'New', 'York', 'ist', 'sollte', 'm
an', "Joe's", 'Pizza', 'probieren.', 'am', 'not', 'a', 'fan', 'of', 'pizza',
'but', 'like', 'the', 'pizza', 'from', 'here.', 'The', 'piece', 'is', 'large'
, 'and', 'with', 'lots', 'of', 'meats', 'on', 'top.', 'hate', 'pizza', 'with'
, 'thick', 'base)', 'love', 'how', 'they', 'also', 'provide', 'chilli', 'fake
s', 'and', 'garlic', 'powder', 'to', 'put', 'on', 'top', 'because', 'love', '
garlic.', 'My', 'favorite', 'pepperoni', 'and', 'fresh', 'mozzarella', 'want'
, 'to', 'eat', 'again.', 'Eating', 'one', 'piece', 'is', 'unfortunate,', 'so'
, 'dont', 'worry,', 'eat', 'two', 'pieces.', 'Large', 'pieces', 'of', 'pizza'
, 'for', 'each', 'slice.', 'There', 'are', 'premade', 'pizzas', 'available',
'and', 'they', 'warm', 'them', 'up', 'before', 'serving', 'them.', "It's", 'g
reat', 'pizza....be', 'sure', 'to', 'give', 'it', 'a', 'try', 'if', 'your', '
in', 'or', 'near', 'Times', 'Square.', 'The', 'pizza', 'got', 'had', 'spinach
,', 'mozzarella', 'and', 'Pepperoni', 'which', 'were', 'so', 'good.', 'Very',
'good', 'place', 'to', 'take', 'a', 'piece', 'of', 'pizza', 'down', 'the', 's
treet.', 'Its', 'a', 'famous', 'place,', 'but', 'the', 'pizza', 'was', 'avera
ge', 'quality.', 'The', 'best', 'pizza', 'in', 'the', 'world', 'Fantastic', '
slice', 'of', 'pie.', 'Perfect', 'balance', 'of', 'cheese,', 'sauce,', 'and',
'dough.', 'Great', 'thin', 'crust', 'style', 'pizza.', 'This', 'definitely',
"isn't", 'the', 'best', 'pizza', 'in', "I'd", 'personally', 'say', "it's", 'j
ust', 'above', 'average', 'for', 'a', 'quick', 'slice.', 'This', 'is', 'no',
'Neapolitan', 'pizza,', 'that', 'is', 'for', 'sure.', 'So,', 'if', "you're",
'looking', 'for', 'a', 'mind', 'blowing', 'authentic', 'Italian', 'slice', 't
his', "isn't", 'it.', 'still', 'think', 'of', 'this', 'pizza', 'How', 'can',
'you', 'complain', 'about', 'cheese', 'pizza', 'when', 'every', 'thing', 'els
```

e', 'around', 'New', 'York', 'City', 'starts', 'at', 'Best', 'slice', 'around
,', 'the', 'line', 'can', 'get', 'quite', 'long', 'but', 'dont', 'be', 'decei
ved', 'this', 'pizza', 'oue', 'is', 'the', 'best', 'youre', 'going', 'to', 'f
ind', 'around', 'the', 'Times', 'Square', 'are.', 'The', 'supreme', 'specialt
y', 'is', 'a', 'must.', 'The', 'drinks', 'selection', 'is', 'quite', 'allrigh
t', 'from', 'juices', 'to', 'sodas.', 'The', 'pizza', 'is', 'really', 'that',
'good.', 'Amazing', 'pizza.', 'Great', 'pizza,', 'not', 'a', 'lot', 'of', 'se
ating', 'but', 'the', 'pizza', 'is', 'worth', 'standing', 'for!', 'Best', 'pi
zza', 'in', 'Thin', 'crust,', 'delicious', 'cheese.', 'Cheese', 'pizza', "won
't", 'dissapoint', 'Great', 'pizza.', 'It', 'was', 'really', 'the', 'best', '
pizza', 'had', 'in', 'New', 'York.', 'Very', 'thin', 'crust', 'and', 'soft',
'dough.', 'famous', 'pizzeria', 'that', 'became', 'a', 'model', 'for', 'the',
'movie', 'Spider-Man.', 'You', 'can', 'taste', 'hot', 'pizza', 'at', 'a', 're
asonable', 'price.', 'Everything', "you'd", 'want', 'from', 'a', 'Pizza', 'Jo
int', 'Quick', 'grab', 'by', 'the', 'slice', 'pizza', 'spot', 'famous', 'in',
'and', 'worth', 'the', 'line.', 'As', 'soon', 'as', 'we', 'arrived', 'in', 't
he', 'city', 'we', 'made', 'a', 'beeline', 'for', 'as', 'is', 'tradition.', '
got', 'the', 'white', 'ricotta', 'and', 'a', 'plain', 'cheese', 'and', 'my',
'partner', 'got', 'pepperoni.', 'Not', 'only', 'was', 'the', 'taste', 'of', '
the', 'pies', 'great,', 'the', 'wait', 'was', 'short', 'even', 'though', 'it'
, 'was', 'busy', 'and', 'the', 'prices', 'were', 'awesome!', 'Best', 'Pizza',
'in', 'New', 'York', 'and', 'the', 'best', 'Pizza', 'on', 'the', 'east', 'coa
st', '', 'Great', 'tasting', 'slices,', 'fast', 'speeds', 'even', 'with', 'so
me', 'of', 'the', 'Times', 'Sq', 'crowd', 'that', 'comes', 'in', 'sometimes.'
, '(Original).', 'Buena', 'pizza.', 'Good', 'pizza', 'Good', 'attention.', 'A
fter', 'trying', 'these', 'pizzas,', 'all', 'my', 'respect', 'to', 'the', 'Am
ericans.', 'Enjoyed', 'the', 'caprese', 'pizza', 'and', 'pepperoni', 'pizza.'
, 'Nothing', 'makes', 'you', 'feel', 'like', 'youre', 'in', 'Ny', 'than', 'a'
, 'slice', 'of', 'Joes', 'However,', 'as', 'soon', 'as', 'you', 'place', 'an'
, 'order,', 'the', 'pizza', 'will', 'shorten', 'quickly.', 'And', 'it', 'was'
, 'so', 'delicious', 'that', 'only', 'had', 'one', 'piece', 'of', 'pizza', 'i
n', 'the', 'photo!', 'If', 'you', 'go', 'to', 'New', 'York,', 'you', 'want',
'to', 'eat', 'another', 'pizza.', 'Great', 'Pizza', 'joint', 'a', 'few', 'blo
cks', 'away', 'from', 'Times', 'Square.', 'What', 'great', 'pizza!', 'The', '
crispy', 'crust', 'and', 'thin', 'slices', 'make', 'for', 'a', 'delicious', '
meal.', 'The', 'white', 'pizza', 'is', 'also', 'very', 'good.', 'The', 'most'
, 'anticipated', 'pizza', 'place', 'near', 'times', 'square', 'where', 'you',
'will', 'find', 'a', 'line', 'of', 'people', 'trying', 'to', 'get', 'a', 'sli
ce', 'during', 'all', 'times', 'of', 'the', 'day.', 'Attend', 'very', 'fast,'
, 'normal', 'pizza,', 'do', 'not', 'think', "it's", 'as', 'great', 'as', 'the
y', 'say', 'Got', 'spinach', 'and', 'mozzarella.', 'Try', 'cheese', 'slice.',
'One', 'of', 'the', 'best', 'New', 'York', 'Style', 'pizzas', 'ever', 'taste.
', 'The', 'pictures', 'in', 'the', 'walls', 'wit', 'a', 'lot', 'of', 'celebri
ty', 'confirm', 'that.', 'With', 'the', 'Covid-19', 'you', 'cant', 'eat', 'in
side', 'but', 'the', 'guys', 'keep', 'it', 'real', 'and', 'make', 'something'
, 'to', 'eat', 'and', 'go', 'outside', 'that', 'was', 'cool.', 'Fresh', 'food
', 'here', 'people', 'is', 'not', 'frozen', 'pizza.', 'Legit', 'the', 'flavor
', 'and', 'consistency', 'went', 'days', 'in', 'a', 'row.', 'Now', "it's", 'n
ot', 'pizza', 'have', 'ever', 'had', 'or', 'anything', 'but', 'its', 'definit
ely', 'up', 'there.', 'The', 'pizza', 'is', 'very', 'We', 'went', 'to', 'eat'

, 'here', 'for', 'the', 'SpiderMan', 'movie.', 'And', 'we', 'liked', 'the', 'pizza', 'a', 'lot', 'The', 'best', 'pizza', "I've", 'had', 'so', 'far,', 'recommend', 'the', 'basil', 'There', 'were', 'a', 'lot', 'of', 'pizza', 'types', 'and', 'it', 'was', 'good', 'because', 'it', 'was', 'not', 'expensive!', 'Their', 'white', 'pizza', 'and', 'supreme', 'pizza', 'is', 'delicious', 'and', 'the', 'quick', 'service', 'is', 'really', 'fast.', 'Great', 'pizza', 'shop', 'on', 'Broadway!', 'Delicious', 'pizza,', 'Maybe', 'the', 'best', 'en', 'recomended', 'cheese', 'pizza!!', 'The', 'pepperoni', 'is', 'also', 'great', 'but', 'the', 'cheese', 'is', 'by', 'far', 'the', 'best', "I've", 'ever', 'had.', 'Delicious', 'pizza!', 'No', 'wonder', 'the', 'place', 'is', 'so', 'famous,', 'they', 'have', 'many', 'kinds', 'of', 'delicious', 'pizza', 'and', '', 'with', 'a', 'quick', 'service.', 'Rich', 'pizza', 'Extremely', 'good', 'pizza.', 'Not', 'very', 'spacious,', 'be', 'prepared', 'to', 'eat', 'standing.', 'But', "it's", 'hard', 'to', 'call', 'the', 'best', 'pizza', 'in', 'New', 'York;)', 'Amazing,', 'quite', 'possibly', 'some', 'of', 'the', 'best', 'pizza', 'in', 'Lots', 'of', 'great', 'options', 'for', 'vegetarians,', 'recommend', 'the', 'plain', 'one,', 'with', 'just', 'cheese', 'and', 'tomato', 'sauce', 'as', 'the', 'one', 'to', 'get', 'followed', 'by', 'one', 'of', 'the', 'white', 'pies.', 'Really', 'awesome', 'pizza.', 'Very', 'good', 'pizza', 'with', 'good', 'muzzarela', 'cheese.', 'like', 'my', 'pizza', 'more', 'crasty', 'but', 'was', 'satisfied.', '(Original).', 'Excelente', 'pizza!', 'Excellent', 'pizza!', 'Heard', 'a', 'lot', 'about', "Joe's", 'Pizza', 'but', 'more', 'than', 'lived', 'up', 'to', 'the', 'hype.', 'Ordered', 'a', 'New', 'York', 'slice', 'and', 'supreme', 'and', 'both', 'were', 'delicious.', 'The', 'base', 'was', 'crispy', 'and', 'thin', 'and', 'the', 'toppings', 'were', 'light', 'and', 'flavorful.', 'Service', 'was', 'quick', 'and', 'easy', 'and', 'we', 'got', 'our', 'Pizza', 'with', 'little', 'fuss.', 'Very', 'good', 'pizza!', 'Service', 'is', 'also', 'great', 'and', 'love', 'that', 'the', 'serve', 'Coke', 'products.', 'Nice', 'ratio', 'of', 'thinness', 'and', 'a', 'little', 'crisp...with', 'good', 'flavor..', 'thing', 'is', 'next', 'day', 'we', 'came', 'back', 'again', 'and', 'the', 'crispness', 'of', 'the', 'slice', 'is', 'not', 'the', 'same', 'anymore..', 'still', 'a', 'must', 'in', 'new', 'york', '(Original).', 'Lekkere', 'en', 'zeer', 'verse', 'pizza!!!!!!', 'Tasty', 'and', 'very', 'fresh', 'pizza', 'The', 'best', 'pizza', 'we', 'have', 'eaten', 'in', 'America!', 'nice', 'thin', 'slice.', 'Very', 'good', 'place', 'to', 'eat', 'the', 'typical', 'pizza.', 'The', 'taste', 'of', 'the', 'pizzas', 'is', 'great.', 'You', 'can', 'eat', 'it', 'standing', 'on', 'one', 'side', 'or', 'ask', 'it', 'to', 'go.', 'El', 'gusto', 'de', 'las', 'pizzas', 'es', 'genial.', 'Another', 'great', 'pizza', 'spot,', 'this', 'on', 'on', 'Broadway.', 'Either', 'way,', 'youll', 'love', 'the', 'pizza.', 'By', 'the', 'slice', 'or', 'a', 'whole', 'pizza', 'to', 'order.', 'Pizza', 'was', 'amazing!', 'Tasty', 'pizza,', 'legendary', 'location.', 'Typical', 'American', 'pizza.', 'personally', 'prefer', 'Italian', 'pizzas.', 'Busy', 'place', '(line', 'moves', 'quickly)', 'but', 'a', 'delicious', 'slice!', 'This', 'is', 'a', 'very', 'good', 'New', 'York', 'Pizza.', "I'm", 'from', 'Boston', 'and', 'my', 'very', 'favorite', 'pizza', 'in', 'the', 'world', 'is', "Regina's", 'in', 'the', 'North', 'end', 'of', 'Boston', 'which', 'is', 'why', "I'm", 'very', 'stingy', 'with', 'five', 'Stars.', 'Too', 'much', 'hype', 'for', 'an', 'average', 'pizza!', 'Umm..', "don't", 'get', 'me', 'wrong,', 'the', 'pizza', 'was', 'ok', 'but', 'for', 'a', 'slice?', 'Fast', 'paced', 'by', 'the', 'slice', 'goodness!', 'One', 'or', 'if', 'not,', 'th

e', 'best', 'pizza', "I've", 'ever', 'had,', 'very', 'picturesque', 'place', 'This', 'pizza', 'really', 'is', 'one', 'of', 'the', 'best', 'pizzas', 'Ive', 'tasted!', 'The', 'pepperoni', 'and', 'white', 'pizza', 'were', 'both', 'very ', 'good', 'and', 'it', 'was', 'quick.', 'Big', 'greasy', 'slices', 'served', 'fast', 'on', 'paper', 'plates.', 'Amazing', 'pizza!', 'Only', 'had', 'one', 'type', 'of', 'pizza', 'to', 'offer.', 'Pizza', 'by', 'the', 'slice', 'or', ' whole', 'pies.', 'One', 'of', 'the', 'Best', 'pizza', 'spots', 'in', 'New', ' York', 'city', 'love', 'this', 'place.', 'really', 'liked', 'this', 'pizza.', 'pizza', 'is', 'one', 'of', 'my', 'favorites.', 'know', 'yummy', 'pizza', 'wh en', 'taste', 'it.', 'the', 'grape', 'soda', '', 'was', 'really', 'good', 'as ', 'well.', "Joe's", 'is', 'a', 'star', 'because', 'it', 'is', 'a', 'top', 'n ew', 'York', 'slice', 'for', 'me.', 'My', 'numbers', 'and', 'require', 'you', 'buy', 'a', 'whole', 'pie', 'so', 'by', 'the', 'slice,', "Joe's", 'is', 'This ', 'is', 'strictly', 'for', 'a', 'classic', 'slice.', "haven't", 'tried', 'an y', 'other', 'slices', 'here.', 'Good', 'pizza,', 'very', 'busy', 'and', 'cro wded', 'but', 'line', 'goes', 'fast.', 'Eleven', 'pm', 'and', "there's", 'a', 'line...this', 'is', 'a', 'good', 'slice', 'of', 'pie!', 'The', 'pizzas', 'ar e', 'good', 'the', 'place', 'is', 'small', 'for', 'so', 'many', 'people', 'Th ey', 'have', 'huge', 'slices', 'of', 'pizza', 'at', 'a', 'great', 'price.', ' Average', 'pizza.', 'personally', 'liked', 'pizza', 'had', 'the', 'day', 'bef ore', 'better.', 'The', 'dough', 'is', 'crispy', 'and', 'the', 'American', 'p izza', 'is', 'delicious.', 'It', 'does', 'always', 'seems', 'to', 'be', 'busy ', 'but', 'things', 'are', 'quick', 'and', 'youll', 'get', 'your', 'pizza', ' slice', 'in', 'no', 'time.', 'Average', 'cost', 'per', 'slice', 'is', 'about' , 'give', 'or', 'take.', 'Best', 'slice', 'ever', 'Mmmmmm', 'could', 'eat', ' all', 'day', 'Great', 'Pizza,', 'Fast', 'Service,', "don't", 'let', 'the', 'l ine', 'scare', 'you,', 'it', 'moves', 'The', 'quintessential', 'New', 'York', 'slice.', 'Perfect', 'crust,', 'simply', 'good', 'pizza.', 'An', 'institution ', 'in', 'times', 'Square', 'for', 'a', 'quick', 'and', 'tasty', 'slice', 'of ', 'pizza', 'so', 'many', 'tastes', 'available', "It's", 'over', 'hyped', 'pi zza', 'for', 'tourists.', "It's", 'a', 'solid', 'slice', 'but', 'nothing', 's pecial.', 'You', 'can', 'easily', 'pay', 'for', 'people', 'for', 'roughly', ' and', 'that', 'including', 'two', 'slices', 'per', 'person', 'with', 'a', 'dr ink.', 'Very', 'nice', 'pizzeria', 'at', 'the', 'heart', 'of', 'new', 'York', 'city.', "don't", 'like', 'pizza', 'very', 'much', 'but', "didn't", 'expect', 'much', 'but', "it's", 'really', 'delicious.', 'Very', 'good', 'pizzas', 'at' , 'a', 'very', 'reasonable', 'price!', '(Original).', 'Super', 'leckere', 'Pi zza', 'und', 'sehr', 'coole', 'Location.', 'Super', 'delicious', 'pizza', 'an d', 'very', 'cool', 'location.', 'If', 'you', 'like', 'pizza', 'you', 'should ', 'definitely', 'visit', "Joe's", 'Pizza.', 'This', 'pizzeria', 'may', 'well ', 'be', 'called', 'an', 'institution', 'in', 'Times', 'Square.', 'Wer', 'Piz za', 'mag', 'sollte', 'unbedingt', "Joe's", 'Pizza', 'einen', 'Besuch', 'abst atten.', 'Diese', 'Pizzeria', 'kann', 'wohl', 'als', 'eine', 'Institution', ' am', 'Times', 'Square', 'bezeichnet', 'werden.', 'Piece', 'of', 'pizza', 'can ', 'eat', 'cheaply', 'The', 'taste', 'seems', 'to', 'be', 'less', 'than', 'th e', 'pizza', 'ate', 'in', 'Sulgangpo', 'got', 'three', 'pies', 'of', 'cheese' , 'pizza', 'four', 'pies', 'of', 'pizza', 'and', 'two', 'pieces', 'of', 'man' , 'Best', 'pizza', 'in', 'the', 'world,', 'flew', 'from', 'Argentina', 'and', 'knew', 'where', 'had', 'to', 'eat.', 'This', 'pizza', 'was', 'so', 'good.', 'Excellent', 'pizza!!!', 'If', 'you', 'are', 'looking', 'for', 'an', 'authent

ic', 'style', 'pizza....this', 'is', 'the', 'place!!', 'Super', 'busy,', 'but
', 'was', 'worth', 'the', 'wait,', 'a', 'real', 'slice', 'of', '', 'The', 'pi
zza', 'itself', 'is', 'not', 'the', 'best', 'but', 'just', 'being', 'in', 'th
at', 'iconic', 'atmosphere,', 'makes', 'the', 'experience', 'phenomenal.', 'G
reat', 'pizza.', "It's", 'cheap,', 'but', 'taste', 'good..', 'thin', 'pizza',
'with', 'a', 'strong', 'flavour..', 'Small', 'busy', 'place,', 'ideal', 'for'
, 'a', 'few', 'portions', 'of', 'tasty', 'pizzas.', 'Four', 'portions', 'of',
'pizza', 'and', 'a', 'bottle', 'of', 'water,', 'for', 'about', 'The', 'walls'
, 'are', 'full', 'of', 'photographs', 'of', 'celebrities', 'who', 'have', 'co
me', 'to', 'the', 'pizzeria.', 'Cuatro', 'porciones', 'de', 'pizza', 'y', 'bo
tella', 'de', 'agua,', 'por', 'unos', 'Las', 'paredes', 'estan', 'llenas', 'd
e', 'fotografas', 'de', 'famosos', 'que', 'han', 'acudido', 'a', 'la', 'pizze
ria.', 'Its', 'a', 'must', 'to', 'stop', 'by', 'and', 'eat', 'a', 'slice', 'O
f', 'pizza', 'in', 'the', 'heart', 'Of', 'The', 'place', 'is', 'small', 'but'
, 'the', 'pizza', 'is', 'totally', 'worthy', 'it', '', 'Best', 'pizza', 'in',
'New', 'York!', 'The', 'pizza', 'was', 'good,', 'the', 'best', 'ever', 'had.'
, 'Perfectly', 'cooked', 'with', 'crispy', 'crusts.', 'It', 'was', 'dollars',
'a', 'slice', 'but', 'the', 'slices', 'were', '2x', 'as', 'big', 'as', 'your'
, 'regular', 'slices.', 'Pair', 'with', 'coke', 'or', 'any', 'other', 'refres
hing', 'drink', 'and', 'enjoy!', '(Original).', 'Excelente', 'pizza!!', 'Exce
llent', 'pizza', 'Great,', 'piping', 'hot', 'pizza', 'that', 'stays', 'hot',
'minutes', 'into', 'eating.', 'The', 'crust', 'is', 'super', 'crisp.', 'The',
'pizza', 'is', 'good!!', 'Great', 'example', 'of', 'nyc', 'pizza', 'no', 'com
plaints,', 'the', 'staff', 'was', 'great,', 'the', 'pizza', 'was', 'amazing!!
', 'Not', 'alot', 'of', 'seating', 'so', 'best', 'of', 'luck', 'lol', 'Awesom
e', 'pizza!', 'Exquisite', 'pizza.', 'They', 'are', 'all', 'very', 'good,', '
especially', 'the', '"White"', 'and', 'the', 'margarita.', 'Typical', 'New',
'York', 'pizza', 'in', 'all', 'its', 'essence.', 'Pizza', 'tpica', 'neoyorqui
na', 'en', 'toda', 'su', 'esencia.', 'The', 'New', 'York', 'pizza', 'as', 'se
en', 'in', 'the', 'movies,', 'huge,', 'steaming', 'and', 'really', 'delicious
.', 'if', 'then', 'add', 'two', 'iced', 'beers', 'taken', 'from', 'the', 'Pha
rmacy', 'in', 'front', '(can', 'of', 'Coors,', 'Colorado', 'lager,', 'from',
'710ml', 'to', '2dollari', 'cadauna)', 'make', 'dinner', 'for', 'the', 'great
est!', 'se', 'poi', 'aggiungete', 'due', 'birre', 'ghiacciate', 'prese', 'dal
', 'Pharmacy', 'di', 'fronte', 'lattina', 'di', 'Coors,', 'lager', 'del', 'Co
lorado,', 'da', '710ml', 'a', '2dollari/cadauna)', 'fate', 'cena', 'alla', 'g
randissima!', 'Delicious', 'New', 'York', 'Style', 'pizza', 'that', 'is', 'de
finitely', 'worth', 'the', 'wait.', 'Brutal', 'pizzas', 'Yummmy', 'pizza', ''
, 'really', 'like', 'it...so', 'fresh', 'and', 'really', 'feel', 'like', 'you
', 'eating', 'italian', 'pizza...', 'Sometimes', 'simplicity', 'is', 'all', '
you', 'need,', 'and', "Joe's", 'Pizza', 'delivers.', 'thin,', 'crispy', 'crus
t,', 'gooey', 'cheese,', 'fresh', '(ever', 'so', 'slightly', 'chunky)', 'toma
to', 'sauce,', 'and', 'savory', 'pepperoni.', 'Delicious', 'pizza.', 'Solid',
'variety.', 'One', 'of', 'the', 'best', 'pizzas', 'in', 'This', 'may', 'be',
'more', 'of', 'a', 'nuanced', 'or', 'critical', 'review', 'simply', 'due', 't
o', 'the', 'fact', 'that', 'expectations', 'run', 'high', 'for', 'New', 'York
', 'pizza', 'in', 'New', 'York.', 'The', 'pizza', "isn't", 'bad,', "don't", '
get', 'me', 'wrong,', "it's", 'just', 'not', 'great.', "don't", 'know', 'how'
, 'to', 'explain', 'it,', 'it', "wasn't", 'bad,', 'it', 'was', 'good', 'pizza
,', 'but', 'something', 'seemed', 'to', 'be', 'missing.', 'Awesome', 'staff,'

, 'very', 'welcoming', 'and', 'you', 'can', 'see', 'all', 'the', 'action', 'g
oing', 'on,', 'the', 'skill', 'behind', 'the', 'best', 'pizza', 'in', 'The',
'pizzas', 'themselves', 'are', 'full', 'of', 'flavour', 'and', 'the', 'base',
'is', 'incredible,', 'some', 'folks', "don't", 'eat', 'the', 'crust', 'but',
'this', 'part', 'is', 'just', 'as', 'special,', 'a', 'little', 'pizza', 'base
', 'dipping', 'sauce', 'would', 'be', 'mega', 'to', 'top', 'it', 'off.', "Don
't", 'stop', 'at', 'one', 'or', 'two', 'slices,', 'arrive', 'hungry', 'and',
'try', 'it', 'all.', 'Delicious', 'pizza!', 'had', 'the', 'cheese', 'and', 't
he', 'caprese.', 'Nice', 'big', 'slices', 'that', 'barely', 'fit', 'on', 'the
', 'little', 'greasy', 'paper', 'plates,', 'and', 'needed', 'to', 'be', 'fold
ed', 'to', 'be', 'eaten,', 'as', 'it', 'should', 'be.', 'With', 'obvious', 'p
epperoni!', 'Para', 'comer', 'al', 'paso', 'y', 'probar', 'un', 'slice', 'de'
, 'esos', 'que', 'vimos', 'en', 'todas', 'las', 'series!', 'Con', 'pepperoni'
, 'obvio!', 'Worst', 'place', 'to', 'eat', 'found', 'a', 'short', 'hair', 'in
', 'my', 'pizza', 'have', 'long', 'hair', 'had', 'to', 'try', 'the', 'famous'
, "Joe's", 'Pizza', 'and', 'was', 'not', 'disappointed.', 'This', 'pizza', 'd
eserves', 'its', 'opinions', 'and', 'even', 'always', 'long', 'queue', 'did',
'not', 'discourage', 'me', 'from', 'going', 'there', 'again', 'and', 'again',
'The', 'pizza', 'slices', 'are', 'just', 'excellent', 'and', 'really', 'cheap
', 'at', 'all.', 'Les', 'slices', 'de', 'pizza', 'sont', 'justes', 'excellent
es', 'et', 'vraiment', 'pas', 'cher', 'du', 'tout.', 'Of', 'course,', 'best',
'pizza', 'ever!', 'Had', 'a', 'great', 'time', 'munching', 'on', 'a', 'Pepper
oni', 'pie.', "It's", 'not', 'designed', 'to', 'be', 'the', 'most', 'gourmet'
, 'pizza,', 'but', 'it', 'was', 'a', 'yummy', 'slice', 'and', 'spending', 'bu
cks', 'to', 'get', 'a', 'whole', 'pizza', 'filling', 'up', 'men', 'is', 'well
', 'worth', 'it.', 'It', 'was', 'good', 'for', 'some', 'late', 'night', 'pizz
a!', 'Delicious', 'pizza,', "I'm", 'coming', 'back', "Joe's", 'Pretty', 'good
', 'pizza,', 'Id', 'recommend', 'to', 'go', 'in', 'the', 'afternoon', 'when',
'the', 'pizza', 'is', 'fresh,', 'closer', 'to', 'dark', 'hours', 'they', 'get
', 'a', 'bit', 'cold,', 'he', 'does', 'reheat', 'in', 'the', 'oven.', 'Hugely
', 'popular', 'for', 'slices.', 'Oh,', "don't", 'forget', 'to', 'try', 'the',
'deep', 'dish', 'slices', 'Pizza', 'is', 'not', 'nice', 'at', 'all.', 'Had',
'pepperoni', 'and', 'white', 'and', 'both', 'were', 'so', 'delicious.', 'Pepp
eroni', 'had', 'a', 'very', 'moderate', 'amount', 'of', 'saltiness,', 'and',
'would', 'say', 'that', 'this', 'is', 'what', 'New', 'York', 'Pizza', 'says',
'about', 'this', 'taste.', 'White', 'has', 'ricotta', 'cheese', 'on', 'top',
'of', 'cheese', 'toppings,', 'which', 'is', 'a', 'crystal', 'of', 'lightness.
', 'Great', 'pizza,', 'truly', 'new', 'York', 'style', 'Excellent', 'pizza',
'from', 'this', 'pizza', 'chain', 'originally', 'from', 'New', 'York!', 'reco
mmend', 'the', 'cheese', 'that', 'is', 'the', 'classic', 'site.', 'Loved', 't
his', 'pizza!', 'Big', 'piece,', 'crispy', 'and', 'delicious!', '(Original).'
, 'Pizzas', 'muy', 'ricas', 'y', 'crujientes.', 'Pizzas', 'very', 'rich', 'an
d', 'crispy.', 'Very', 'little', 'variety', 'of', 'pizzas', 'the', 'truth,',
'and', 'the', 'flavors', 'leave', 'a', 'lot', 'to', 'be', 'desired.', 'You',
'can', 'get', 'better', 'pizza', 'at', 'most', 'local', 'suburban', 'places',
'across', 'the', 'country.', 'Great', 'pizza', 'the', 'best', 'of', 'my', 'li
fe', 'fresh', 'basil', 'and', 'mozzy', 'fraiche', 'yummie', 'Proper', 'Slice!
', 'can', 'still', 'smell', 'the', 'delicious', 'Pepperoni!', 'Thin,', 'crisp
', 'base', 'with', 'classic', 'toppings.', 'Hot', 'and', 'tasty', '(though',
'personally', 'prefer', 'neapolitan', 'crusts).', 'They', 'should', 'have', '

written', 'explanations', 'of', 'the', 'different', 'toppings.', 'Very', 'yummy', 'pizza.', 'Best', 'pizza', 'spot', 'in', 'Decent', 'pizza,', 'very', 'busy.', 'must', 'classic', 'thin', 'crust', 'New', 'York', 'style', 'pizza!', 'Even', 'if', "you're", 'not', 'a', 'pizza', 'fan,', "Joe's", 'Pizza', 'will', 'make', 'you', 'think', 'again.', 'Pizza', 'is', 'life.', 'Then', "there's", "Joe's", 'pizza,', 'a', 'bite', 'of', 'heaven', 'in', 'city', 'Of', 'course', "couldn't", 'be', 'in', 'New', 'York', 'and', 'not', 'have', 'at', 'least', 'one', 'night', 'of', 'pizza', 'for', 'dinner.', 'Staying', 'in', 'Midtown', 'had', 'to', 'Google', 'the', 'best', 'pizza', 'place', 'around', 'and', "Joe's", 'Pizza', 'was', 'the', 'top', 'of', 'the', 'list', 'Bar', 'None.', 'got', 'a', 'slice', 'of', 'pepperoni', 'and', 'my', 'mom', 'got', 'a', 'slice', 'of', 'cheese.', 'Not', 'going', 'to', 'say', "it's", 'the', 'best', 'pizza', 'ever', 'had', 'in', 'my', 'life...', 'Not', 'to', 'mention', 'there', 'was', 'a', 'picture', 'of', 'Carrie', 'Bradshaw', 'on', 'the', 'wall', 'and', 'that', 'immediately', 'had', 'me', 'excited.', 'would', 'recommend', 'as', 'the', 'pizza', 'was', 'fresh..', 'yummy...', 'And', 'cheap!', 'Nothing', 'but', 'hype', 'pizza', "wasn't", 'that', 'great', '(Original).', 'Pizza', 'eccezionale', 'anche', 'a', 'New', 'York,', '', 'da', 'non', 'crederci...vedevamo', 'sempre', 'una', 'lunga', 'coda', 'che', 'non', 'potevano', 'non', 'provarla...very', 'good!!', 'Outstanding', 'pizza,', 'even', 'in', 'New', 'York,', 'you', "can't", 'believe', 'it', 'we', 'always', 'saw', 'a', 'long', 'queue', 'that', "couldn't", 'fail', 'to', 'prove', 'it', 'very', 'good', 'Super', 'popular', 'pizza', 'place', 'in', 'New', 'York!', 'Exquisite', 'pizza', 'The', 'prices', 'for', 'portions', 'are', 'based', 'on', 'Tastes,', 'average', 'of', 'dollars.', 'Exquisita', 'pizza', '', 'Los', 'precios', 'por', 'porciones', 'son', 'en', 'funcion', 'de', 'Los', 'gustos,', 'promedio', 'de', 'dolares.', 'Amazing', 'pizza!', 'Perfect', 'place', 'for', 'a', 'quick', 'and', 'delicious', 'slice!', 'Best', 'pizza!', 'Fresh', 'tomato', 'sauce,', 'nice', 'thin', 'crust,', 'pretty', 'good', 'ingredients.', 'One', 'of', 'NYCs', 'best', 'slices.', 'One', 'of', 'the', 'strongest', 'pizza', 'in', 'New', 'York.', 'Very', 'tasty', 'and', 'satiated.', 'And', 'ate', 'it', 'more', 'than', 'once.', 'But', 'the', 'staff', 'unfortunately', 'not', 'serviced.', 'Because', 'to', 'whom', 'did', 'you', 'ask', 'if', 'the', 'meat', 'was', 'pork', 'or', 'not?', 'The', 'best', 'New', 'York', 'slice', 'during', 'my', 'visit.', 'The', 'crust,', 'the', 'sauce,', 'and', 'the', 'cheese', 'was', 'everything', 'had', 'hoped', 'for.', 'Great', 'pizza.', 'Super', 'nice', 'staff', 'and', 'wonderful', 'pizza!', 'This', 'is', 'the', 'gold', 'standard', 'for', 'the', 'revered', 'style', 'pizza.', 'Definitley', 'a', 'proper', 'place', 'to', 'have', 'a', 'style', 'pizza!', 'Compare', 'to', 'the', 'pizza', 'in', 'here,', 'Sbarro', 'that', 'kind', 'of', 'things', 'are', 'rubbish.', "It's", 'a', 'typical', 'New', 'York', 'style', 'pizza', 'with', 'a', 'thin', 'dough', 'and', 'a', 'light,', 'slightly', 'painted', 'topping.', "It's", 'not', 'a', 'classic', 'pizza', 'baked', 'in', 'a', 'charcoal', 'oven,', 'but', 'considering', 'the', 'cost', 'performance,', 'it', 'might', 'be', 'one', 'of', 'the', 'best', 'pizza', 'places.', 'The', 'dough', 'is', 'different', 'for', 'pepperoni', 'pizza', 'and', 'basil', 'pizza.', "It's", 'a', 'pizza', 'that', 'care', 'about', 'that', 'much.', 'Excelente', 'pizzeria.', '', 'a', 'must', 'visit.', 'Pizza', 'is', 'great,', 'prices', 'are', '"normal', 'considering', 'this', 'is', 'New', 'York', 'Pizza', 'is', 'always', 'great,', 'but', 'nothing', 'special', 'can', 'say', 'about', 'this', 'one.', 'Delicious', 'pizza', 'we', 'went', 'twice', 'o

n', 'our', 'visit', 'to', 'New', 'York.', 'The', 'most', 'amazing', 'pizza',
'my', 'husband', 'and', 'have', 'ever', 'ate', 'like', 'Ill', 'never', 'be',
'able', 'to', 'pizza', 'the', 'same', 'again.', 'We', 'went', 'to', 'a', 'cou
ple', 'pizza', 'places', 'and', 'Joes', 'pizza', 'was', 'absolutely', 'the',
'best', 'had', 'the', 'cheese', 'and', 'the', 'spinach', 'and', 'both', 'were
', 'just', 'perfect', 'my', 'husbands', 'favorite', 'was', 'the', 'pepp', 'an
d', 'cheese.', 'My', 'husband', 'is', 'already', 'planning', 'to', 'try', 'an
d', 'make', 'this', 'pizza', 'since', 'we', 'live', 'in', 'Ohio.', 'Tried', '
the', 'cheese,', 'pepp,', 'and', 'caprese.', 'Really', 'good', 'pizza,', "did
n't", 'expect', 'it.', 'The', 'pizza', 'is', 'so', 'yummy', 'we', 'had', 'to'
, 'go', 'back', 'a', 'second', 'night', 'WoooooW.', 'You', 'have', 'to', 'try
', 'it', 'its', 'really', 'one', 'of', 'the', 'best', 'pizza', 'you', 'will',
'eat', 'ever', '(Original).', 'Uma', 'das', 'melhores', 'pizzas', 'de', 'Manh
attan,', 'a', 'melhor', 'da', 'Broadway.', 'One', 'of', 'the', 'best', 'pizza
s', 'in', 'Manhattan,', 'the', 'best', 'on', 'Broadway.', 'Our', 'first', 'Ne
w', 'York', 'pizza', 'and', 'boy', 'was', 'it', 'got', 'the', 'pepperoni', 'a
nd', 'my', 'boyfriend', 'got', 'the', 'margarita.', 'Both', 'delicious.', "Jo
e's", 'is', 'a', 'machine', 'that', 'processes', 'one', 'thing:', 'some', 'of
', 'the', 'best', 'pizza', 'slices', 'in', 'town.', 'The', 'pizza', 'from', "
Joe's", 'is', 'unarguably', 'one', 'of', 'the', 'most', 'delicious', 'thin',
'crust', 'slices', "you'll", 'ever', 'have.', 'Great', 'pizza,', 'quick', 'se
rvice', 'even', 'when', 'full,', 'friendly', 'staff', 'The', 'New', 'York', '
style', 'pizza', 'is', 'truly', 'worth', 'it..', 'Especially', 'you', 'may',
'have', 'slices', 'instead', 'of', 'ordering', 'a', 'complete', 'pizza', 'if'
, 'you', 'are', 'alone', '(Original).', 'Pizza', 'muito', 'boa.', 'Pizza', 'v
ery', 'good.', 'Yummy', 'Great', 'style', 'pizza.', 'Slices', 'are', 'huge.',
'Only', 'downside', 'was', 'that', 'the', 'pizza', 'base', 'was', 'slightly',
'burnt', 'Best', 'pizza', 'in', 'Pizza', 'tasted', 'absolutely', 'amazing.',
'The', 'slices', 'are', 'huge', 'and', 'fresh.', 'One', 'slice', 'could', 'be
', 'enough', 'for', 'one', 'person.', 'Margherita', 'is', 'my', 'favourite.',
'Amazing', 'pizza!!!!!', 'best', 'pizza', 'in', 'nyc', 'a', 'must', 'have', '
wile', 'there', 'For', 'my', 'first', 'slice', 'of', '', 'it', 'was', 'defini
tely', 'worth', 'the', 'walk.', 'Maybe', 'my', 'favorite', 'slice', 'ever.',
'am', 'a', 'pizza', 'chef.', "I've", 'been', 'to', 'many', 'pizzerias', 'in',
'New', 'York.', 'Please', 'enjoy', 'pepperoni.', 'Great', 'slices,', 'a', 'pe
rfect', 'balance', 'of', 'cheese,', 'sauce', 'and', 'the', 'perfect', 'flexib
le', 'but', 'crunchy', 'base.', 'Do', 'not', 'miss', 'this', 'place', 'if', '
you', 'want', 'to', 'taste', 'the', 'authentic', 'pizza', 'slice.', 'The', 'p
lace', 'is', 'very', 'small', 'and', 'there', 'may', 'not', 'be', 'space', 't
o', 'sit,', 'but', 'the', 'pizza', 'is', 'super', 'fresh', 'and', 'delicious.
', 'The', 'slices', 'are', 'big', 'enough', 'to', 'be', 'full', 'with', 'just
', 'one', 'slice', '(at', 'least', 'for', 'me).', 'They', 'had', 'a', 'select
ion', 'of', 'several', 'pizzas', 'and', 'as', 'a', 'vegetarian,', 'chose', 's
pinach', 'pizza,', 'which', 'was', 'super', 'yummy!', 'Even', 'if', "you're",
'long,', "you're", 'out.', 'It', 'is', 'delicious', 'if', 'we', 'eat', 'witho
ut', 'expectation', 'feel', 'its', 'a', 'hyped', 'place', 'for', 'pizzas', 'b
ut', 'still', 'better', 'than', 'few', 'pizzas', 'have', 'had', 'and', 'okayi
sh', 'if', 'compared', 'with', 'few', 'pizzas', 'have', 'had', 'till', 'now.'
, 'was', 'disappointed', 'as', 'a', '1st', 'time', 'patron', 'of', 'this', 'p
izza', 'joint.', 'As', 'you', 'can', 'see', 'from', 'the', 'photo,', 'it', 'w

```
as', 'burnt', 'in', 'all', 'the', 'wrong', 'places', 'and', 'the', 'crust', '
was', 'very', 'course', 'tough...as', 'if', 'they', 'used', 'old', 'dough.',
'They', 'delivered', 'burnt', 'pizza', 'without', 'the', 'toppings', 'request
ed', 'and', 'it', 'was', 'room', 'temperature.', 'The', 'pizza', 'is', 'very'
, 'rich', 'but', 'do', 'not', 'think', 'it', 'is', 'the', 'best', 'in', 'thou
ght', 'the', 'pizza', 'was', 'good', 'but', 'not', 'really', '"all"', 'that'."
, 'Omg', 'great', 'pizza.', 'Especially', 'the', 'ricotta', 'pizza', '', 'The
', 'pizza', 'was', 'thin', 'and', 'crunchy', 'and', 'cheesey!', 'The', 'marga
rita', 'pizza', 'was', 'also', 'fresh', 'and', 'delicious!', 'The', 'perfect'
, 'quintessential', 'New', 'York', 'Pizza', 'slice', 'The', 'best', 'pizza',
'in', 'the', 'city.', 'Great', 'pizza.', 'Therere', 'limited', 'seatings', 'i
n', 'store', 'so', 'we', 'have', 'the', 'pizza', 'at', 'Bryant', 'Park.', 'Co
mvel', 'no', 'one', 'makes', 'pizza', 'like', 'in', 'Brazil', 'Delicious', 'p
izzas', 'Pepperoni,', 'Supreme', 'and', 'Quattro', 'Cheese', 'Pizza', 'was',
'a', 'bit', 'spicy,', 'but', 'Supreme', 'and', 'Quattro', 'Pizza', 'were', 'v
ery', 'delicious.', 'Even', 'though', 'ate', 'at', 'the', 'hotel', 'minutes',
'after', 'packing,', 'the', 'dough', 'was', 'chewy', 'and', 'the', 'cheese',
'was', 'alive.', 'Personally,', 'only', 'ate', 'Supreme', 'Pizza,', 'but', 'f
elt', 'that', "Greenwich's", "Joe's", 'Pizza', 'was', 'more', 'delicious.', '
Still', 'the', 'best', 'pizza', 'place', 'to', 'eat', 'late', 'at', 'dawn.',
'However', 'while', 'the', 'pizza', 'was', 'extremely', 'delicious,', "wouldn
't", 'find', 'myself', 'spending', 'dollars', 'for', 'a', 'slice', 'of', 'pep
peroni.', 'find', 'spending', 'dollars', 'for', 'three', 'slices', 'rather',
'unnecessary', 'Instant', 'hot', 'pizzas', 'perform', 'well', 'to', 'grow', '
thin', 'dough.', 'Aside', 'from', 'being', 'a', 'bit', 'edgy,', 'the', 'pizza
', 'was', 'delicious', 'and', 'it', 'was', 'worth', 'the', '15-minute', 'queu
e.', 'Aparte', 'de', 'que', 'fue', 'un', 'poco', 'borde,', 'la', 'pizza', 'es
taba', 'buensima', 'y', 'mereci', 'la', 'pena', 'hacer', 'cola', 'de', 'min.'
, 'One', 'of', 'the', 'best', 'pizza', "you'll", 'put', 'in', 'your', 'mouth.
', 'People,', 'locals', 'and', 'travelers', 'alike,', 'will', 'stand', 'in',
'the', 'rain', 'to', 'get', 'a', 'slice', 'here.', 'got', 'a', 'sausage', 'pi
e', 'cuz', "it's", 'real', 'sliced', 'sausage,', 'not', 'crumbled', 'sausage'
, 'like', 'in', 'other', 'cities.', 'They', 'usually', 'have', 'plain,', 'pep
peroni,', 'fresh', 'mozz,', 'caprese,', 'white,', 'and', 'a', 'mixed', 'toppi
ng', 'pizza', 'really', 'available.']
```

Lowercase Convertion

The below code snippet performs lowercase conversion on two lists (alpha_reviews_one
and alpha_reviews_two) and stores the results in lowercase_doc1 and lowercase_doc2
respectively.

```
In [73]:   #lowercase conversion
           lowercase_doc1 = [word.lower() for word in alpha_reviews_one]
           lowercase_doc2 = [word.lower() for word in alpha_reviews_two]
           print(lowercase_doc1)
```

['awesome', 'pizza', '(original).', 'hammer', 'pizza', 'und', 'sie', 'schmeck
t', 'einfach', 'kstlich.', 'hammer', 'pizza', 'and', 'it', 'just', 'tastes',
'delicious.', 'also', 'wenn', 'man', 'in', 'new', 'york', 'ist', 'sollte', 'm
an', "joe's", 'pizza', 'probieren.', 'am', 'not', 'a', 'fan', 'of', 'pizza',
'but', 'like', 'the', 'pizza', 'from', 'here.', 'the', 'piece', 'is', 'large'
, 'and', 'with', 'lots', 'of', 'meats', 'on', 'top.', 'hate', 'pizza', 'with'
, 'thick', 'base)', 'love', 'how', 'they', 'also', 'provide', 'chilli', 'fake
s', 'and', 'garlic', 'powder', 'to', 'put', 'on', 'top', 'because', 'love', '
garlic.', 'my', 'favorite', 'pepperoni', 'and', 'fresh', 'mozzarella', 'want'
, 'to', 'eat', 'again.', 'eating', 'one', 'piece', 'is', 'unfortunate,', 'so'
, 'dont', 'worry,', 'eat', 'two', 'pieces.', 'large', 'pieces', 'of', 'pizza'
, 'for', 'each', 'slice.', 'there', 'are', 'premade', 'pizzas', 'available',
'and', 'they', 'warm', 'them', 'up', 'before', 'serving', 'them.', "it's", 'g
reat', 'pizza....be', 'sure', 'to', 'give', 'it', 'a', 'try', 'if', 'your', '
in', 'or', 'near', 'times', 'square.', 'the', 'pizza', 'got', 'had', 'spinach
,', 'mozzarella', 'and', 'pepperoni', 'which', 'were', 'so', 'good.', 'very',
'good', 'place', 'to', 'take', 'a', 'piece', 'of', 'pizza', 'down', 'the', 's
treet.', 'its', 'a', 'famous', 'place,', 'but', 'the', 'pizza', 'was', 'avera
ge', 'quality.', 'the', 'best', 'pizza', 'in', 'the', 'world', 'fantastic', '
slice', 'of', 'pie.', 'perfect', 'balance', 'of', 'cheese,', 'sauce,', 'and',
'dough.', 'great', 'thin', 'crust', 'style', 'pizza.', 'this', 'definitely',
"isn't", 'the', 'best', 'pizza', 'in', "i'd", 'personally', 'say', "it's", 'j
ust', 'above', 'average', 'for', 'a', 'quick', 'slice.', 'this', 'is', 'no',
'neapolitan', 'pizza,', 'that', 'is', 'for', 'sure.', 'so,', 'if', "you're",
'looking', 'for', 'a', 'mind', 'blowing', 'authentic', 'italian', 'slice', 't
his', "isn't", 'it.', 'still', 'think', 'of', 'this', 'pizza', 'how', 'can',
'you', 'complain', 'about', 'cheese', 'pizza', 'when', 'every', 'thing', 'els
e', 'around', 'new', 'york', 'city', 'starts', 'at', 'best', 'slice', 'around
,', 'the', 'line', 'can', 'get', 'quite', 'long', 'but', 'dont', 'be', 'decei
ved', 'this', 'pizza', 'oue', 'is', 'the', 'best', 'youre', 'going', 'to', 'f
ind', 'around', 'the', 'times', 'square', 'are.', 'the', 'supreme', 'specialt
y', 'is', 'a', 'must.', 'the', 'drinks', 'selection', 'is', 'quite', 'allrigh
t', 'from', 'juices', 'to', 'sodas.', 'the', 'pizza', 'is', 'really', 'that',
'good.', 'amazing', 'pizza.', 'great', 'pizza,', 'not', 'a', 'lot', 'of', 'se
ating', 'but', 'the', 'pizza', 'is', 'worth', 'standing', 'for!', 'best', 'pi
zza', 'in', 'thin', 'crust,', 'delicious', 'cheese.', 'cheese', 'pizza', "won
't", 'dissapoint', 'great', 'pizza.', 'it', 'was', 'really', 'the', 'best', '
pizza', 'had', 'in', 'new', 'york.', 'very', 'thin', 'crust', 'and', 'soft',
'dough.', 'famous', 'pizzeria', 'that', 'became', 'a', 'model', 'for', 'the',
'movie', 'spider-man.', 'you', 'can', 'taste', 'hot', 'pizza', 'at', 'a', 're
asonable', 'price.', 'everything', "you'd", 'want', 'from', 'a', 'pizza', 'jo
int', 'quick', 'grab', 'by', 'the', 'slice', 'pizza', 'spot', 'famous', 'in',
'and', 'worth', 'the', 'line.', 'as', 'soon', 'as', 'we', 'arrived', 'in', 't
he', 'city', 'we', 'made', 'a', 'beeline', 'for', 'as', 'is', 'tradition.', '
got', 'the', 'white', 'ricotta', 'and', 'a', 'plain', 'cheese', 'and', 'my',
'partner', 'got', 'pepperoni.', 'not', 'only', 'was', 'the', 'taste', 'of', '
the', 'pies', 'great,', 'the', 'wait', 'was', 'short', 'even', 'though', 'it'
, 'was', 'busy', 'and', 'the', 'prices', 'were', 'awesome!', 'best', 'pizza',
'in', 'new', 'york', 'and', 'the', 'best', 'pizza', 'on', 'the', 'east', 'coa
st', '', 'great', 'tasting', 'slices,', 'fast', 'speeds', 'even', 'with', 'so

me', 'of', 'the', 'times', 'sq', 'crowd', 'that', 'comes', 'in', 'sometimes.'
, '(original).', 'buena', 'pizza.', 'good', 'pizza', 'good', 'attention.', 'a
fter', 'trying', 'these', 'pizzas,', 'all', 'my', 'respect', 'to', 'the', 'am
ericans.', 'enjoyed', 'the', 'caprese', 'pizza', 'and', 'pepperoni', 'pizza.'
, 'nothing', 'makes', 'you', 'feel', 'like', 'youre', 'in', 'ny', 'than', 'a'
, 'slice', 'of', 'joes', 'however,', 'as', 'soon', 'as', 'you', 'place', 'an'
, 'order,', 'the', 'pizza', 'will', 'shorten', 'quickly.', 'and', 'it', 'was'
, 'so', 'delicious', 'that', 'only', 'had', 'one', 'piece', 'of', 'pizza', 'i
n', 'the', 'photo!', 'if', 'you', 'go', 'to', 'new', 'york,', 'you', 'want',
'to', 'eat', 'another', 'pizza.', 'great', 'pizza', 'joint', 'a', 'few', 'blo
cks', 'away', 'from', 'times', 'square.', 'what', 'great', 'pizza!', 'the', '
crispy', 'crust', 'and', 'thin', 'slices', 'make', 'for', 'a', 'delicious', '
meal.', 'the', 'white', 'pizza', 'is', 'also', 'very', 'good.', 'the', 'most'
, 'anticipated', 'pizza', 'place', 'near', 'times', 'square', 'where', 'you',
'will', 'find', 'a', 'line', 'of', 'people', 'trying', 'to', 'get', 'a', 'sli
ce', 'during', 'all', 'times', 'of', 'the', 'day.', 'attend', 'very', 'fast,'
, 'normal', 'pizza,', 'do', 'not', 'think', "it's", 'as', 'great', 'as', 'the
y', 'say', 'got', 'spinach', 'and', 'mozzarella.', 'try', 'cheese', 'slice.',
'one', 'of', 'the', 'best', 'new', 'york', 'style', 'pizzas', 'ever', 'taste.
', 'the', 'pictures', 'in', 'the', 'walls', 'wit', 'a', 'lot', 'of', 'celebri
ty', 'confirm', 'that.', 'with', 'the', 'covid-19', 'you', 'cant', 'eat', 'in
side', 'but', 'the', 'guys', 'keep', 'it', 'real', 'and', 'make', 'something'
, 'to', 'eat', 'and', 'go', 'outside', 'that', 'was', 'cool.', 'fresh', 'food
', 'here', 'people', 'is', 'not', 'frozen', 'pizza.', 'legit', 'the', 'flavor
', 'and', 'consistency', 'went', 'days', 'in', 'a', 'row.', 'now', "it's", 'n
ot', 'pizza', 'have', 'ever', 'had', 'or', 'anything', 'but', 'its', 'definit
ely', 'up', 'there.', 'the', 'pizza', 'is', 'very', 'we', 'went', 'to', 'eat'
, 'here', 'for', 'the', 'spiderman', 'movie.', 'and', 'we', 'liked', 'the', '
pizza', 'a', 'lot', 'the', 'best', 'pizza', "i've", 'had', 'so', 'far,', 'rec
ommend', 'the', 'basil', 'there', 'were', 'a', 'lot', 'of', 'pizza', 'types',
'and', 'it', 'was', 'good', 'because', 'it', 'was', 'not', 'expensive!', 'the
ir', 'white', 'pizza', 'and', 'supreme', 'pizza', 'is', 'delicious', 'and', '
the', 'quick', 'service', 'is', 'really', 'fast.', 'great', 'pizza', 'shop',
'on', 'broadway!', 'delicious', 'pizza,', 'maybe', 'the', 'best', 'en', 'reco
mended', 'cheese', 'pizza!!', 'the', 'pepperoni', 'is', 'also', 'great', 'but
', 'the', 'cheese', 'is', 'by', 'far', 'the', 'best', "i've", 'ever', 'had.',
'delicious', 'pizza!', 'no', 'wonder', 'the', 'place', 'is', 'so', 'famous,',
'they', 'have', 'many', 'kinds', 'of', 'delicious', 'pizza', 'and', '', 'with
', 'a', 'quick', 'service.', 'rich', 'pizza', 'extremely', 'good', 'pizza.',
'not', 'very', 'spacious,', 'be', 'prepared', 'to', 'eat', 'standing.', 'but'
, "it's", 'hard', 'to', 'call', 'the', 'best', 'pizza', 'in', 'new', 'york;)'
, 'amazing,', 'quite', 'possibly', 'some', 'of', 'the', 'best', 'pizza', 'in'
, 'lots', 'of', 'great', 'options', 'for', 'vegetarians,', 'recommend', 'the'
, 'plain', 'one,', 'with', 'just', 'cheese', 'and', 'tomato', 'sauce', 'as',
'the', 'one', 'to', 'get', 'followed', 'by', 'one', 'of', 'the', 'white', 'pi
es.', 'really', 'awesome', 'pizza.', 'very', 'good', 'pizza', 'with', 'good',
'muzzarela', 'cheese.', 'like', 'my', 'pizza', 'more', 'crasty', 'but', 'was'
, 'satisfied.', '(original).', 'excelente', 'pizza!', 'excellent', 'pizza!',
'heard', 'a', 'lot', 'about', "joe's", 'pizza', 'but', 'more', 'than', 'lived
', 'up', 'to', 'the', 'hype.', 'ordered', 'a', 'new', 'york', 'slice', 'and',

'supreme', 'and', 'both', 'were', 'delicious.', 'the', 'base', 'was', 'crispy
', 'and', 'thin', 'and', 'the', 'toppings', 'were', 'light', 'and', 'flavorfu
l.', 'service', 'was', 'quick', 'and', 'easy', 'and', 'we', 'got', 'our', 'pi
zza', 'with', 'little', 'fuss.', 'very', 'good', 'pizza!', 'service', 'is', '
also', 'great', 'and', 'love', 'that', 'the', 'serve', 'coke', 'products.', '
nice', 'ratio', 'of', 'thinness', 'and', 'a', 'little', 'crisp...with', 'good
', 'flavor..', 'thing', 'is', 'next', 'day', 'we', 'came', 'back', 'again', '
and', 'the', 'crispness', 'of', 'the', 'slice', 'is', 'not', 'the', 'same', '
anymore..', 'still', 'a', 'must', 'in', 'new', 'york', '(original).', 'lekker
e', 'en', 'zeer', 'verse', 'pizza!!!!!!', 'tasty', 'and', 'very', 'fresh', 'p
izza', 'the', 'best', 'pizza', 'we', 'have', 'eaten', 'in', 'america!', 'nice
', 'thin', 'slice.', 'very', 'good', 'place', 'to', 'eat', 'the', 'typical',
'pizza.', 'the', 'taste', 'of', 'the', 'pizzas', 'is', 'great.', 'you', 'can'
, 'eat', 'it', 'standing', 'on', 'one', 'side', 'or', 'ask', 'it', 'to', 'go.
', 'el', 'gusto', 'de', 'las', 'pizzas', 'es', 'genial.', 'another', 'great',
'pizza', 'spot,', 'this', 'on', 'on', 'broadway.', 'either', 'way,', 'youll',
'love', 'the', 'pizza.', 'by', 'the', 'slice', 'or', 'a', 'whole', 'pizza', '
to', 'order.', 'pizza', 'was', 'amazing!', 'tasty', 'pizza,', 'legendary', 'l
ocation.', 'typical', 'american', 'pizza.', 'personally', 'prefer', 'italian'
, 'pizzas.', 'busy', 'place', '(line', 'moves', 'quickly)', 'but', 'a', 'deli
cious', 'slice!', 'this', 'is', 'a', 'very', 'good', 'new', 'york', 'pizza.',
"i'm", 'from', 'boston', 'and', 'my', 'very', 'favorite', 'pizza', 'in', 'the
', 'world', 'is', "regina's", 'in', 'the', 'north', 'end', 'of', 'boston', 'w
hich', 'is', 'why', "i'm", 'very', 'stingy', 'with', 'five', 'stars.', 'too',
'much', 'hype', 'for', 'an', 'average', 'pizza!', 'umm..', "don't", 'get', 'm
e', 'wrong,', 'the', 'pizza', 'was', 'ok', 'but', 'for', 'a', 'slice?', 'fast
', 'paced', 'by', 'the', 'slice', 'goodness!', 'one', 'or', 'if', 'not,', 'th
e', 'best', 'pizza', "i've", 'ever', 'had,', 'very', 'picturesque', 'place',
'this', 'pizza', 'really', 'is', 'one', 'of', 'the', 'best', 'pizzas', 'ive',
'tasted!', 'the', 'pepperoni', 'and', 'white', 'pizza', 'were', 'both', 'very
', 'good', 'and', 'it', 'was', 'quick.', 'big', 'greasy', 'slices', 'served',
'fast', 'on', 'paper', 'plates.', 'amazing', 'pizza!', 'only', 'had', 'one',
'type', 'of', 'pizza', 'to', 'offer.', 'pizza', 'by', 'the', 'slice', 'or', '
whole', 'pies.', 'one', 'of', 'the', 'best', 'pizza', 'spots', 'in', 'new', '
york', 'city', 'love', 'this', 'place.', 'really', 'liked', 'this', 'pizza.',
'pizza', 'is', 'one', 'of', 'my', 'favorites.', 'know', 'yummy', 'pizza', 'wh
en', 'taste', 'it.', 'the', 'grape', 'soda', '', 'was', 'really', 'good', 'as
', 'well.', "joe's", 'is', 'a', 'star', 'because', 'it', 'is', 'a', 'top', 'n
ew', 'york', 'slice', 'for', 'me.', 'my', 'numbers', 'and', 'require', 'you',
'buy', 'a', 'whole', 'pie', 'so', 'by', 'the', 'slice,', "joe's", 'is', 'this
', 'is', 'strictly', 'for', 'a', 'classic', 'slice.', "haven't", 'tried', 'an
y', 'other', 'slices', 'here.', 'good', 'pizza,', 'very', 'busy', 'and', 'cro
wded', 'but', 'line', 'goes', 'fast.', 'eleven', 'pm', 'and', "there's", 'a',
'line...this', 'is', 'a', 'good', 'slice', 'of', 'pie!', 'the', 'pizzas', 'ar
e', 'good', 'the', 'place', 'is', 'small', 'for', 'so', 'many', 'people', 'th
ey', 'have', 'huge', 'slices', 'of', 'pizza', 'at', 'a', 'great', 'price.', '
average', 'pizza.', 'personally', 'liked', 'pizza', 'had', 'the', 'day', 'bef
ore', 'better.', 'the', 'dough', 'is', 'crispy', 'and', 'the', 'american', 'p
izza', 'is', 'delicious.', 'it', 'does', 'always', 'seems', 'to', 'be', 'busy
', 'but', 'things', 'are', 'quick', 'and', 'youll', 'get', 'your', 'pizza', '

slice', 'in', 'no', 'time.', 'average', 'cost', 'per', 'slice', 'is', 'about'
, 'give', 'or', 'take.', 'best', 'slice', 'ever', 'mmmmmm', 'could', 'eat', '
all', 'day', 'great', 'pizza,', 'fast', 'service,', "don't", 'let', 'the', 'l
ine', 'scare', 'you,', 'it', 'moves', 'the', 'quintessential', 'new', 'york',
'slice.', 'perfect', 'crust,', 'simply', 'good', 'pizza.', 'an', 'institution
', 'in', 'times', 'square', 'for', 'a', 'quick', 'and', 'tasty', 'slice', 'of
', 'pizza', 'so', 'many', 'tastes', 'available', "it's", 'over', 'hyped', 'pi
zza', 'for', 'tourists.', "it's", 'a', 'solid', 'slice', 'but', 'nothing', 's
pecial.', 'you', 'can', 'easily', 'pay', 'for', 'people', 'for', 'roughly', '
and', 'that', 'including', 'two', 'slices', 'per', 'person', 'with', 'a', 'dr
ink.', 'very', 'nice', 'pizzeria', 'at', 'the', 'heart', 'of', 'new', 'york',
'city.', "don't", 'like', 'pizza', 'very', 'much', 'but', "didn't", 'expect',
'much', 'but', "it's", 'really', 'delicious.', 'very', 'good', 'pizzas', 'at'
, 'a', 'very', 'reasonable', 'price!', '(original).', 'super', 'leckere', 'pi
zza', 'und', 'sehr', 'coole', 'location.', 'super', 'delicious', 'pizza', 'an
d', 'very', 'cool', 'location.', 'if', 'you', 'like', 'pizza', 'you', 'should
', 'definitely', 'visit', "joe's", 'pizza.', 'this', 'pizzeria', 'may', 'well
', 'be', 'called', 'an', 'institution', 'in', 'times', 'square.', 'wer', 'piz
za', 'mag', 'sollte', 'unbedingt', "joe's", 'pizza', 'einen', 'besuch', 'abst
atten.', 'diese', 'pizzeria', 'kann', 'wohl', 'als', 'eine', 'institution', '
am', 'times', 'square', 'bezeichnet', 'werden.', 'piece', 'of', 'pizza', 'can
', 'eat', 'cheaply', 'the', 'taste', 'seems', 'to', 'be', 'less', 'than', 'th
e', 'pizza', 'ate', 'in', 'sulgangpo', 'got', 'three', 'pies', 'of', 'cheese'
, 'pizza', 'four', 'pies', 'of', 'pizza', 'and', 'two', 'pieces', 'of', 'man'
, 'best', 'pizza', 'in', 'the', 'world,', 'flew', 'from', 'argentina', 'and',
'knew', 'where', 'had', 'to', 'eat.', 'this', 'pizza', 'was', 'so', 'good.',
'excellent', 'pizza!!!', 'if', 'you', 'are', 'looking', 'for', 'an', 'authent
ic', 'style', 'pizza....this', 'is', 'the', 'place!!', 'super', 'busy,', 'but
', 'was', 'worth', 'the', 'wait,', 'a', 'real', 'slice', 'of', '', 'the', 'pi
zza', 'itself', 'is', 'not', 'the', 'best', 'but', 'just', 'being', 'in', 'th
at', 'iconic', 'atmosphere,', 'makes', 'the', 'experience', 'phenomenal.', 'g
reat', 'pizza.', "it's", 'cheap,', 'but', 'taste', 'good..', 'thin', 'pizza',
'with', 'a', 'strong', 'flavour..', 'small', 'busy', 'place,', 'ideal', 'for'
, 'a', 'few', 'portions', 'of', 'tasty', 'pizzas.', 'four', 'portions', 'of',
'pizza', 'and', 'a', 'bottle', 'of', 'water,', 'for', 'about', 'the', 'walls'
, 'are', 'full', 'of', 'photographs', 'of', 'celebrities', 'who', 'have', 'co
me', 'to', 'the', 'pizzeria.', 'cuatro', 'porciones', 'de', 'pizza', 'y', 'bo
tella', 'de', 'agua,', 'por', 'unos', 'las', 'paredes', 'estan', 'llenas', 'd
e', 'fotografas', 'de', 'famosos', 'que', 'han', 'acudido', 'a', 'la', 'pizze
ria.', 'its', 'a', 'must', 'to', 'stop', 'by', 'and', 'eat', 'a', 'slice', 'o
f', 'pizza', 'in', 'the', 'heart', 'of', 'the', 'place', 'is', 'small', 'but'
, 'the', 'pizza', 'is', 'totally', 'worthy', 'it', '', 'best', 'pizza', 'in',
'new', 'york!', 'the', 'pizza', 'was', 'good,', 'the', 'best', 'ever', 'had.'
, 'perfectly', 'cooked', 'with', 'crispy', 'crusts.', 'it', 'was', 'dollars',
'a', 'slice', 'but', 'the', 'slices', 'were', '2x', 'as', 'big', 'as', 'your'
, 'regular', 'slices.', 'pair', 'with', 'coke', 'or', 'any', 'other', 'refres
hing', 'drink', 'and', 'enjoy!', '(original).', 'excelente', 'pizza!!', 'exce
llent', 'pizza', 'great,', 'piping', 'hot', 'pizza', 'that', 'stays', 'hot',
'minutes', 'into', 'eating.', 'the', 'crust', 'is', 'super', 'crisp.', 'the',
'pizza', 'is', 'good!!', 'great', 'example', 'of', 'nyc', 'pizza', 'no', 'com

plaints,', 'the', 'staff', 'was', 'great,', 'the', 'pizza', 'was', 'amazing!!
', 'not', 'alot', 'of', 'seating', 'so', 'best', 'of', 'luck', 'lol', 'awesom
e', 'pizza!', 'exquisite', 'pizza.', 'they', 'are', 'all', 'very', 'good,', '
especially', 'the', '"white"', 'and', 'the', 'margarita.', 'typical', 'new',
'york', 'pizza', 'in', 'all', 'its', 'essence.', 'pizza', 'tpica', 'neoyorqui
na', 'en', 'toda', 'su', 'esencia.', 'the', 'new', 'york', 'pizza', 'as', 'se
en', 'in', 'the', 'movies,', 'huge,', 'steaming', 'and', 'really', 'delicious
.', 'if', 'then', 'add', 'two', 'iced', 'beers', 'taken', 'from', 'the', 'pha
rmacy', 'in', 'front', '(can', 'of', 'coors,', 'colorado', 'lager,', 'from',
'710ml', 'to', '2dollari', 'cadauna)', 'make', 'dinner', 'for', 'the', 'great
est!', 'se', 'poi', 'aggiungete', 'due', 'birre', 'ghiacciate', 'prese', 'dal
', 'pharmacy', 'di', 'fronte', 'lattina', 'di', 'coors,', 'lager', 'del', 'co
lorado,', 'da', '710ml', 'a', '2dollari/cadauna)', 'fate', 'cena', 'alla', 'g
randissima!', 'delicious', 'new', 'york', 'style', 'pizza', 'that', 'is', 'de
finitely', 'worth', 'the', 'wait.', 'brutal', 'pizzas', 'yummmy', 'pizza', ''
, 'really', 'like', 'it...so', 'fresh', 'and', 'really', 'feel', 'like', 'you
', 'eating', 'italian', 'pizza...', 'sometimes', 'simplicity', 'is', 'all', '
you', 'need,', 'and', "joe's", 'pizza', 'delivers.', 'thin,', 'crispy', 'crus
t,', 'gooey', 'cheese,', 'fresh', '(ever', 'so', 'slightly', 'chunky)', 'toma
to', 'sauce,', 'and', 'savory', 'pepperoni.', 'delicious', 'pizza.', 'solid',
'variety.', 'one', 'of', 'the', 'best', 'pizzas', 'in', 'this', 'may', 'be',
'more', 'of', 'a', 'nuanced', 'or', 'critical', 'review', 'simply', 'due', 't
o', 'the', 'fact', 'that', 'expectations', 'run', 'high', 'for', 'new', 'york
', 'pizza', 'in', 'new', 'york.', 'the', 'pizza', "isn't", 'bad,', "don't", '
get', 'me', 'wrong,', "it's", 'just', 'not', 'great.', "don't", 'know', 'how'
, 'to', 'explain', 'it,', 'it', "wasn't", 'bad,', 'it', 'was', 'good', 'pizza
,', 'but', 'something', 'seemed', 'to', 'be', 'missing.', 'awesome', 'staff,'
, 'very', 'welcoming', 'and', 'you', 'can', 'see', 'all', 'the', 'action', 'g
oing', 'on,', 'the', 'skill', 'behind', 'the', 'best', 'pizza', 'in', 'the',
'pizzas', 'themselves', 'are', 'full', 'of', 'flavour', 'and', 'the', 'base',
'is', 'incredible,', 'some', 'folks', "don't", 'eat', 'the', 'crust', 'but',
'this', 'part', 'is', 'just', 'as', 'special,', 'a', 'little', 'pizza', 'base
', 'dipping', 'sauce', 'would', 'be', 'mega', 'to', 'top', 'it', 'off.', "don
't", 'stop', 'at', 'one', 'or', 'two', 'slices,', 'arrive', 'hungry', 'and',
'try', 'it', 'all.', 'delicious', 'pizza!', 'had', 'the', 'cheese', 'and', 't
he', 'caprese.', 'nice', 'big', 'slices', 'that', 'barely', 'fit', 'on', 'the
', 'little', 'greasy', 'paper', 'plates,', 'and', 'needed', 'to', 'be', 'fold
ed', 'to', 'be', 'eaten,', 'as', 'it', 'should', 'be.', 'with', 'obvious', 'p
epperoni!', 'para', 'comer', 'al', 'paso', 'y', 'probar', 'un', 'slice', 'de'
, 'esos', 'que', 'vimos', 'en', 'todas', 'las', 'series!', 'con', 'pepperoni'
, 'obvio!', 'worst', 'place', 'to', 'eat', 'found', 'a', 'short', 'hair', 'in
', 'my', 'pizza', 'have', 'long', 'hair', 'had', 'to', 'try', 'the', 'famous'
, "joe's", 'pizza', 'and', 'was', 'not', 'disappointed.', 'this', 'pizza', 'd
eserves', 'its', 'opinions', 'and', 'even', 'always', 'long', 'queue', 'did',
'not', 'discourage', 'me', 'from', 'going', 'there', 'again', 'and', 'again',
'the', 'pizza', 'slices', 'are', 'just', 'excellent', 'and', 'really', 'cheap
', 'at', 'all.', 'les', 'slices', 'de', 'pizza', 'sont', 'justes', 'excellent
es', 'et', 'vraiment', 'pas', 'cher', 'du', 'tout.', 'of', 'course,', 'best',
'pizza', 'ever!', 'had', 'a', 'great', 'time', 'munching', 'on', 'a', 'pepper
oni', 'pie.', "it's", 'not', 'designed', 'to', 'be', 'the', 'most', 'gourmet'

, 'pizza,', 'but', 'it', 'was', 'a', 'yummy', 'slice', 'and', 'spending', 'bucks', 'to', 'get', 'a', 'whole', 'pizza', 'filling', 'up', 'men', 'is', 'well', 'worth', 'it.', 'it', 'was', 'good', 'for', 'some', 'late', 'night', 'pizza!', 'delicious', 'pizza,', "i'm", 'coming', 'back', "joe's", 'pretty', 'good', 'pizza,', 'id', 'recommend', 'to', 'go', 'in', 'the', 'afternoon', 'when', 'the', 'pizza', 'is', 'fresh,', 'closer', 'to', 'dark', 'hours', 'they', 'get', 'a', 'bit', 'cold,', 'he', 'does', 'reheat', 'in', 'the', 'oven.', 'hugely', 'popular', 'for', 'slices.', 'oh,', "don't", 'forget', 'to', 'try', 'the', 'deep', 'dish', 'slices', 'pizza', 'is', 'not', 'nice', 'at', 'all.', 'had', 'pepperoni', 'and', 'white', 'and', 'both', 'were', 'so', 'delicious.', 'pepperoni', 'had', 'a', 'very', 'moderate', 'amount', 'of', 'saltiness,', 'and', 'would', 'say', 'that', 'this', 'is', 'what', 'new', 'york', 'pizza', 'says', 'about', 'this', 'taste.', 'white', 'has', 'ricotta', 'cheese', 'on', 'top', 'of', 'cheese', 'toppings,', 'which', 'is', 'a', 'crystal', 'of', 'lightness.', 'great', 'pizza,', 'truly', 'new', 'york', 'style', 'excellent', 'pizza', 'from', 'this', 'pizza', 'chain', 'originally', 'from', 'new', 'york!', 'recommend', 'the', 'cheese', 'that', 'is', 'the', 'classic', 'site.', 'loved', 'this', 'pizza!', 'big', 'piece,', 'crispy', 'and', 'delicious!', '(original).', 'pizzas', 'muy', 'ricas', 'y', 'crujientes.', 'pizzas', 'very', 'rich', 'and', 'crispy.', 'very', 'little', 'variety', 'of', 'pizzas', 'the', 'truth,', 'and', 'the', 'flavors', 'leave', 'a', 'lot', 'to', 'be', 'desired.', 'you', 'can', 'get', 'better', 'pizza', 'at', 'most', 'local', 'suburban', 'places', 'across', 'the', 'country.', 'great', 'pizza', 'the', 'best', 'of', 'my', 'life', 'fresh', 'basil', 'and', 'mozzy', 'fraiche', 'yummie', 'proper', 'slice!', 'can', 'still', 'smell', 'the', 'delicious', 'pepperoni!', 'thin,', 'crisp', 'base', 'with', 'classic', 'toppings.', 'hot', 'and', 'tasty', '(though', 'personally', 'prefer', 'neapolitan', 'crusts).', 'they', 'should', 'have', 'written', 'explanations', 'of', 'the', 'different', 'toppings.', 'very', 'yummy', 'pizza.', 'best', 'pizza', 'spot', 'in', 'decent', 'pizza,', 'very', 'busy.', 'must', 'classic', 'thin', 'crust', 'new', 'york', 'style', 'pizza!', 'even', 'if', "you're", 'not', 'a', 'pizza', 'fan,', "joe's", 'pizza', 'will', 'make', 'you', 'think', 'again.', 'pizza', 'is', 'life.', 'then', "there's", "joe's", 'pizza,', 'a', 'bite', 'of', 'heaven', 'in', 'city', 'of', 'course', "couldn't", 'be', 'in', 'new', 'york', 'and', 'not', 'have', 'at', 'least', 'one', 'night', 'of', 'pizza', 'for', 'dinner.', 'staying', 'in', 'midtown', 'had', 'to', 'google', 'the', 'best', 'pizza', 'place', 'around', 'and', "joe's", 'pizza', 'was', 'the', 'top', 'of', 'the', 'list', 'bar', 'none.', 'got', 'a', 'slice', 'of', 'pepperoni', 'and', 'my', 'mom', 'got', 'a', 'slice', 'of', 'cheese.', 'not', 'going', 'to', 'say', "it's", 'the', 'best', 'pizza', 'ever', 'had', 'in', 'my', 'life...', 'not', 'to', 'mention', 'there', 'was', 'a', 'picture', 'of', 'carrie', 'bradshaw', 'on', 'the', 'wall', 'and', 'that', 'immediately', 'had', 'me', 'excited.', 'would', 'recommend', 'as', 'the', 'pizza', 'was', 'fresh..', 'yummy...', 'and', 'cheap!', 'nothing', 'but', 'hype', 'pizza', "wasn't", 'that', 'great', '(original).', 'pizza', 'eccezionale', ',', 'anche', 'a', 'new', 'york,', '', 'da', 'non', 'crederci...vedevamo', 'sempre', 'una', 'lunga', 'coda', 'che', 'non', 'potevano', 'non', 'provarla...very', 'good!!', 'outstanding', 'pizza,', 'even', 'in', 'new', 'york,', 'you', "can't", 'believe', 'it', 'we', 'always', 'saw', 'a', 'long', 'queue', 'that', "couldn't", 'fail', 'to', 'prove', 'it', 'very', 'good', 'super', 'popular', 'pizza', 'place', 'in', 'new', 'york!', 'exquisite', 'pizza', 'the', 'price

s', 'for', 'portions', 'are', 'based', 'on', 'tastes,', 'average', 'of', 'dollars.', 'exquisita', 'pizza', '', 'los', 'precios', 'por', 'porciones', 'son', 'en', 'funcion', 'de', 'los', 'gustos,', 'promedio', 'de', 'dolares.', 'amazing', 'pizza!', 'perfect', 'place', 'for', 'a', 'quick', 'and', 'delicious', 'slice!', 'best', 'pizza!', 'fresh', 'tomato', 'sauce,', 'nice', 'thin', 'crust,', 'pretty', 'good', 'ingredients.', 'one', 'of', 'nycs', 'best', 'slices.', 'one', 'of', 'the', 'strongest', 'pizza', 'in', 'new', 'york.', 'very', 'tasty', 'and', 'satiated.', 'and', 'ate', 'it', 'more', 'than', 'once.', 'but', 'the', 'staff', 'unfortunately', 'not', 'serviced.', 'because', 'to', 'whom', 'did', 'you', 'ask', 'if', 'the', 'meat', 'was', 'pork', 'or', 'not?', 'the', 'best', 'new', 'york', 'slice', 'during', 'my', 'visit.', 'the', 'crust,', 'the', 'sauce,', 'and', 'the', 'cheese', 'was', 'everything', 'had', 'hoped', 'for.', 'great', 'pizza.', 'super', 'nice', 'staff', 'and', 'wonderful', 'pizza!', 'this', 'is', 'the', 'gold', 'standard', 'for', 'the', 'revered', 'style', 'pizza.', 'definitley', 'a', 'proper', 'place', 'to', 'have', 'a', 'style', 'pizza!', 'compare', 'to', 'the', 'pizza', 'in', 'here,', 'sbarro', 'that', 'kind', 'of', 'things', 'are', 'rubbish.', "it's", 'a', 'typical', 'new', 'york', 'style', 'pizza', 'with', 'a', 'thin', 'dough', 'and', 'a', 'light,', 'slightly', 'painted', 'topping.', "it's", 'not', 'a', 'classic', 'pizza', 'baked', 'in', 'a', 'charcoal', 'oven,', 'but', 'considering', 'the', 'cost', 'performance,', 'it', 'might', 'be', 'one', 'of', 'the', 'best', 'pizza', 'places.', 'the', 'dough', 'is', 'different', 'for', 'pepperoni', 'pizza', 'and', 'basil', 'pizza.', "it's", 'a', 'pizza', 'that', 'care', 'about', 'that', 'much.', 'excelente', 'pizzeria.', '', 'a', 'must', 'visit.', 'pizza', 'is', 'great,', 'prices', 'are', '"normal', 'considering', 'this', 'is', 'new', 'york', 'pizza', 'is', 'always', 'great,', 'but', 'nothing', 'special', 'can', 'say', 'about', 'this', 'one.', 'delicious', 'pizza', 'we', 'went', 'twice', 'on', 'our', 'visit', 'to', 'new', 'york.', 'the', 'most', 'amazing', 'pizza', 'my', 'husband', 'and', 'have', 'ever', 'ate', 'like', 'ill', 'never', 'be', 'able', 'to', 'pizza', 'the', 'same', 'again.', 'we', 'went', 'to', 'a', 'couple', 'pizza', 'places', 'and', 'joes', 'pizza', 'was', 'absolutely', 'the', 'best', 'had', 'the', 'cheese', 'and', 'the', 'spinach', 'and', 'both', 'were', 'just', 'perfect', 'my', 'husbands', 'favorite', 'was', 'the', 'pepp', 'and', 'cheese.', 'my', 'husband', 'is', 'already', 'planning', 'to', 'try', 'and', 'make', 'this', 'pizza', 'since', 'we', 'live', 'in', 'ohio.', 'tried', 'the', 'cheese,', 'pepp,', 'and', 'caprese.', 'really', 'good', 'pizza,', "didn't", 'expect', 'it.', 'the', 'pizza', 'is', 'so', 'yummy', 'we', 'had', 'to', 'go', 'back', 'a', 'second', 'night', 'wooooow.', 'you', 'have', 'to', 'try', 'it', 'its', 'really', 'one', 'of', 'the', 'best', 'pizza', 'you', 'will', 'eat', 'ever', '(original).', 'uma', 'das', 'melhores', 'pizzas', 'de', 'manhattan,', 'a', 'melhor', 'da', 'broadway.', 'one', 'of', 'the', 'best', 'pizzas', 'in', 'manhattan,', 'the', 'best', 'on', 'broadway.', 'our', 'first', 'new', 'york', 'pizza', 'and', 'boy', 'was', 'it', 'got', 'the', 'pepperoni', 'and', 'my', 'boyfriend', 'got', 'the', 'margarita.', 'both', 'delicious.', "joe's", 'is', 'a', 'machine', 'that', 'processes', 'one', 'thing:', 'some', 'of', 'the', 'best', 'pizza', 'slices', 'in', 'town.', 'the', 'pizza', 'from', "joe's", 'is', 'unarguably', 'one', 'of', 'the', 'most', 'delicious', 'thin', 'crust', 'slices', "you'll", 'ever', 'have.', 'great', 'pizza,', 'quick', 'service', 'even', 'when', 'full,', 'friendly', 'staff', 'the', 'new', 'york', 'style', 'pizza', 'is', 'truly', 'worth', 'it..', 'especially', 'you', 'may',

'have', 'slices', 'instead', 'of', 'ordering', 'a', 'complete', 'pizza', 'if', 'you', 'are', 'alone', '(original).', 'pizza', 'muito', 'boa.', 'pizza', 'very', 'good.', 'yummy', 'great', 'style', 'pizza.', 'slices', 'are', 'huge.', 'only', 'downside', 'was', 'that', 'the', 'pizza', 'base', 'was', 'slightly', 'burnt', 'best', 'pizza', 'in', 'pizza', 'tasted', 'absolutely', 'amazing.', 'the', 'slices', 'are', 'huge', 'and', 'fresh.', 'one', 'slice', 'could', 'be', 'enough', 'for', 'one', 'person.', 'margherita', 'is', 'my', 'favourite.', 'amazing', 'pizza!!!!!', 'best', 'pizza', 'in', 'nyc', 'a', 'must', 'have', 'wile', 'there', 'for', 'my', 'first', 'slice', 'of', '', 'it', 'was', 'definitely', 'worth', 'the', 'walk.', 'maybe', 'my', 'favorite', 'slice', 'ever.', 'am', 'a', 'pizza', 'chef.', "i've", 'been', 'to', 'many', 'pizzerias', 'in', 'new', 'york.', 'please', 'enjoy', 'pepperoni.', 'great', 'slices,', 'a', 'perfect', 'balance', 'of', 'cheese,', 'sauce', 'and', 'the', 'perfect', 'flexible', 'but', 'crunchy', 'base.', 'do', 'not', 'miss', 'this', 'place', 'if', 'you', 'want', 'to', 'taste', 'the', 'authentic', 'pizza', 'slice.', 'the', 'place', 'is', 'very', 'small', 'and', 'there', 'may', 'not', 'be', 'space', 'to', 'sit,', 'but', 'the', 'pizza', 'is', 'super', 'fresh', 'and', 'delicious.', 'the', 'slices', 'are', 'big', 'enough', 'to', 'be', 'full', 'with', 'just', 'one', 'slice', '(at', 'least', 'for', 'me).', 'they', 'had', 'a', 'selection', 'of', 'several', 'pizzas', 'and', 'as', 'a', 'vegetarian,', 'chose', 'spinach', 'pizza,', 'which', 'was', 'super', 'yummy!', 'even', 'if', "you're", 'long,', "you're", 'out.', 'it', 'is', 'delicious', 'if', 'we', 'eat', 'without', 'expectation', 'feel', 'its', 'a', 'hyped', 'place', 'for', 'pizzas', 'but', 'still', 'better', 'than', 'few', 'pizzas', 'have', 'had', 'and', 'okayish', 'if', 'compared', 'with', 'few', 'pizzas', 'have', 'had', 'till', 'now.', 'was', 'disappointed', 'as', 'a', '1st', 'time', 'patron', 'of', 'this', 'pizza', 'joint.', 'as', 'you', 'can', 'see', 'from', 'the', 'photo,', 'it', 'was', 'burnt', 'in', 'all', 'the', 'wrong', 'places', 'and', 'the', 'crust', 'was', 'very', 'course', 'tough...as', 'if', 'they', 'used', 'old', 'dough.', 'they', 'delivered', 'burnt', 'pizza', 'without', 'the', 'toppings', 'requested', 'and', 'it', 'was', 'room', 'temperature.', 'the', 'pizza', 'is', 'very', 'rich', 'but', 'do', 'not', 'think', 'it', 'is', 'the', 'best', 'in', 'thought', 'the', 'pizza', 'was', 'good', 'but', 'not', 'really', "'all", "that'.", 'omg', 'great', 'pizza.', 'especially', 'the', 'ricotta', 'pizza', '', 'the', 'pizza', 'was', 'thin', 'and', 'crunchy', 'and', 'cheesey!', 'the', 'margarita', 'pizza', 'was', 'also', 'fresh', 'and', 'delicious!', 'the', 'perfect', 'quintessential', 'new', 'york', 'pizza', 'slice', 'the', 'best', 'pizza', 'in', 'the', 'city.', 'great', 'pizza.', 'therere', 'limited', 'seatings', 'in', 'store', 'so', 'we', 'have', 'the', 'pizza', 'at', 'bryant', 'park.', 'comvel', 'no', 'one', 'makes', 'pizza', 'like', 'in', 'brazil', 'delicious', 'pizzas', 'pepperoni,', 'supreme', 'and', 'quattro', 'cheese', 'pizza', 'was', 'a', 'bit', 'spicy,', 'but', 'supreme', 'and', 'quattro', 'pizza', 'were', 'very', 'delicious.', 'even', 'though', 'ate', 'at', 'the', 'hotel', 'minutes', 'after', 'packing,', 'the', 'dough', 'was', 'chewy', 'and', 'the', 'cheese', 'was', 'alive.', 'personally,', 'only', 'ate', 'supreme', 'pizza,', 'but', 'felt', 'that', "greenwich's", "joe's", 'pizza', 'was', 'more', 'delicious.', 'still', 'the', 'best', 'pizza', 'place', 'to', 'eat', 'late', 'at', 'dawn.', 'however', 'while', 'the', 'pizza', 'was', 'extremely', 'delicious,', "wouldn't", 'find', 'myself', 'spending', 'dollars', 'for', 'a', 'slice', 'of', 'pepperoni.', 'find', 'spending', 'dollars', 'for', 'three', 'slices', 'rather',

'unnecessary', 'instant', 'hot', 'pizzas', 'perform', 'well', 'to', 'grow', 'thin', 'dough.', 'aside', 'from', 'being', 'a', 'bit', 'edgy,', 'the', 'pizza', 'was', 'delicious', 'and', 'it', 'was', 'worth', 'the', '15-minute', 'queue.', 'aparte', 'de', 'que', 'fue', 'un', 'poco', 'borde,', 'la', 'pizza', 'estaba', 'buensima', 'y', 'mereci', 'la', 'pena', 'hacer', 'cola', 'de', 'min.', 'one', 'of', 'the', 'best', 'pizza', "you'll", 'put', 'in', 'your', 'mouth.', 'people,', 'locals', 'and', 'travelers', 'alike,', 'will', 'stand', 'in', 'the', 'rain', 'to', 'get', 'a', 'slice', 'here.', 'got', 'a', 'sausage', 'pie', 'cuz', "it's", 'real', 'sliced', 'sausage,', 'not', 'crumbled', 'sausage', 'like', 'in', 'other', 'cities.', 'they', 'usually', 'have', 'plain,', 'pepperoni,', 'fresh', 'mozz,', 'caprese,', 'white,', 'and', 'a', 'mixed', 'topping', 'pizza', 'really', 'available.']

stopwords

The code snippet performs stop word removal on the lowercase_doc1 and lowercase_doc2 lists using the combined list of stopwords.

```
In [74]:  nltkstopwords = nltk.corpus.stopwords.words('english')
          morestopwords = ['a','.','b','y',"'s","'d","'ll","'t","'m","'re","'ve",'coul
          finalstopwords= nltkstopwords+morestopwords
          stopped_doc1= [w for w in lowercase_doc1 if not w in finalstopwords]
          stopped_doc2= [w for w in lowercase_doc2 if not w in finalstopwords]
          print(stopped_doc1[:50])
          print(stopped_doc2[:50])
```

['awesome', 'pizza', '(original).', 'hammer', 'pizza', 'und', 'sie', 'schmeckt', 'einfach', 'kstlich.', 'hammer', 'pizza', 'tastes', 'delicious.', 'also', 'wenn', 'man', 'new', 'york', 'ist', 'sollte', 'man', "joe's", 'pizza', 'probieren.', 'fan', 'pizza', 'like', 'pizza', 'here.', 'piece', 'large', 'lots', 'meats', 'top.', 'hate', 'pizza', 'thick', 'base)', 'love', 'also', 'provide', 'chilli', 'fakes', 'garlic', 'powder', 'put', 'top', 'love', 'garlic.']
['one', 'best', 'pizza', 'spots', 'new', 'york', 'city', 'love', 'place.', 'feel', 'hyped', 'place', 'pizzas', 'still', 'better', 'pizzas', 'okayish', 'compared', 'pizzas', 'till', 'now.', 'great', 'pizza.', 'therere', 'limited', 'seatings', 'store', 'pizza', 'bryant', 'park.', 'try', 'cheese', 'slice.', 'white', 'pizza', 'supreme', 'pizza', 'delicious', 'quick', 'service', 'really', 'fast.', 'best', 'slice', 'ever', 'mmmmmm', 'eat', 'day', 'another', 'great']

```
In [62]:  nltk.download('wordnet')
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]     /Users/kashyapchaganti/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

Out[62]: True

lemmatization

The code snippet performs lemmatization on the stopped_doc1 and stopped_doc2 lists using the WordNet lemmatizer from NLTK.

```python
In [75]:  from nltk.stem import WordNetLemmatizer
          lemmatizer = WordNetLemmatizer()
          lemmatized_words_doc1 = [lemmatizer.lemmatize(word) for word in stopped_doc1
          lemmatized_words_doc2 = [lemmatizer.lemmatize(word) for word in stopped_doc2
          print(lemmatized_words_doc1[:50])
          print(lemmatized_words_doc2[:50])
```

```
['awesome', 'pizza', '(original).', 'hammer', 'pizza', 'und', 'sie', 'schmeck
t', 'einfach', 'kstlich.', 'hammer', 'pizza', 'taste', 'delicious.', 'also',
'wenn', 'man', 'new', 'york', 'ist', 'sollte', 'man', "joe's", 'pizza', 'prob
ieren.', 'fan', 'pizza', 'like', 'pizza', 'here.', 'piece', 'large', 'lot', '
meat', 'top.', 'hate', 'pizza', 'thick', 'base)', 'love', 'also', 'provide',
'chilli', 'fake', 'garlic', 'powder', 'put', 'top', 'love', 'garlic.']
['one', 'best', 'pizza', 'spot', 'new', 'york', 'city', 'love', 'place.', 'fe
el', 'hyped', 'place', 'pizza', 'still', 'better', 'pizza', 'okayish', 'compa
red', 'pizza', 'till', 'now.', 'great', 'pizza.', 'therere', 'limited', 'seat
ing', 'store', 'pizza', 'bryant', 'park.', 'try', 'cheese', 'slice.', 'white'
, 'pizza', 'supreme', 'pizza', 'delicious', 'quick', 'service', 'really', 'fa
st.', 'best', 'slice', 'ever', 'mmmmmm', 'eat', 'day', 'another', 'great']
```

Final Stop Words

The provided code defines a function filter_stopwords that filters words from a given list based on several conditions as given below (I used alpha filter and stop words filter along with symbols)

```python
In [81]:  def filter_stopwords(words_list, stopwords_list, symbols_list):
              filtered_words = []
              for word in words_list:
                  if word.casefold() not in stopwords_list and word not in symbols_lis
                      filtered_words.append(word)
              return filtered_words
          symbols = "!\"#$%&()*+-./:;<=>?@[\]^_`{|}~\n'"
          finalwords_doc1= filter_stopwords(lemmatized_words_doc1,finalstopwords,symbc
          finalwords_doc2= filter_stopwords(lemmatized_words_doc2,finalstopwords,symbc
          print(finalwords_doc1[:50])
          print(finalwords_doc2[:50])
```

```
['awesome', 'pizza', 'hammer', 'pizza', 'und', 'sie', 'schmeckt', 'einfach',
'kstlich.', 'hammer', 'pizza', 'taste', 'delicious.', 'also', 'wenn', 'man',
'new', 'york', 'ist', 'sollte', 'man', "joe's", 'pizza', 'probieren.', 'fan',
'pizza', 'like', 'pizza', 'here.', 'piece', 'large', 'lot', 'meat', 'top.', '
hate', 'pizza', 'thick', 'base)', 'love', 'also', 'provide', 'chilli', 'fake'
, 'garlic', 'powder', 'put', 'top', 'love', 'garlic.', 'favorite']
['one', 'best', 'pizza', 'spot', 'new', 'york', 'city', 'love', 'place.', 'fe
el', 'hyped', 'place', 'pizza', 'still', 'better', 'pizza', 'okayish', 'compa
red', 'pizza', 'till', 'now.', 'great', 'pizza.', 'therere', 'limited', 'seat
ing', 'store', 'pizza', 'bryant', 'park.', 'try', 'cheese', 'slice.', 'white'
, 'pizza', 'supreme', 'pizza', 'delicious', 'quick', 'service', 'really', 'fa
st.', 'best', 'slice', 'ever', 'mmmmmm', 'eat', 'day', 'another', 'great']
```

Top50 Frquency

The code defines a function named frequencyTop50 that calculates the top 50 words by
frequency in a given document. Here's a summary of what the code does:

1. The function takes a single argument doc, which is the document for which we want
   to calculate the top 50 words by frequency.
2. The variable length is assigned the length of the document.
3. The nltk.FreqDist() function is used to create a frequency distribution of the words
   in the document. It counts the occurrences of each word.
4. The freqdist object is then used to create a new frequency distribution newfreqdist
   where the frequency of each word is normalized by dividing it by the length of the
   document.
5. The newfreqdist.items() returns a list of word-frequency pairs, which is then sorted
   based on the frequency in descending order using the sorted() function and a
   lambda function as the key.
6. The sorted list of word-frequency pairs is then sliced to retrieve the top 50 pairs
   using the expression sortedFreq[:50].
7. The resulting list of the top 50 word-frequency pairs is returned by the function.

You can call this function by passing the document as an argument, and it will return a
list of the top 50 words by frequency, normalized by the length of the document.

```python
In [64]:  def frequencyTop50(doc):
              length = len(doc)
              freqdist = nltk.FreqDist(doc)
              newfreqdist = {word: freq / length for word, freq in freqdist.items()}
              sortedFreq = sorted(newfreqdist.items(), key=lambda x: x[1], reverse=Tru
              return sortedFreq[:50]
```

```
In [82]:   print("First Doc")
           print()
           print(frequencyTop50(finalwords_doc1))
           print()
           print("Second Doc")
           print()
           print(frequencyTop50(finalwords_doc2))
```

First Doc

```
[('pizza', 0.09221772379667116), ('slice', 0.02249212775528565), ('best', 0.0
206927575348628), ('new', 0.016644174538911382), ('great', 0.0121457489878542
51), ('pizza.', 0.012145748987854251), ('one', 0.011695906432748537), ('york'
, 0.011246063877642825), ('good', 0.011246063877642825), ('place', 0.00944669
3657219974), ('delicious', 0.00899685110211426), ('pizza,', 0.008547008547008
548), ('eat', 0.007647323436797121), ('cheese', 0.007197480881691408), ('real
ly', 0.007197480881691408), ('pizza!', 0.007197480881691408), ("joe's", 0.006
747638326585695), ('thin', 0.005847953216374269), ('pepperoni', 0.00539811066
1268556), ('de', 0.005398110661268556), ('time', 0.004948268106162843), ('got
', 0.004948268106162843), ('style', 0.004948268106162843), ('delicious.', 0.0
04449842555105713), ('like', 0.00449842555105713), ('fresh', 0.004498425551057
13), ('get', 0.00449842555105713), ('ever', 0.00449842555105713), ('taste', 0
.004048582995951417), ('quick', 0.004048582995951417), ('lot', 0.003598740440
845704), ('crust', 0.003598740440845704), ('worth', 0.003598740440845704), ('
even', 0.003598740440845704), ('make', 0.003598740440845704), ('super', 0.003
598740440845704), ('piece', 0.003148897885739991), ('slice.', 0.0031488978857
39991), ('try', 0.003148897885739991), ('perfect', 0.003148897885739991), ('w
hite', 0.003148897885739991), ('nice', 0.003148897885739991), ('also', 0.0026
99055330634278), ('average', 0.002699055330634278), ('say', 0.002699055330634
278), ('supreme', 0.002699055330634278), ('crispy', 0.002699055330634278), ('
tasty', 0.002699055330634278), ('la', 0.002699055330634278), ('love', 0.00224
9212775528565)]
```

Second Doc

```
[('pizza', 0.09197475202885483), ('slice', 0.02254283137962128), ('best', 0.0
20739404869251576), ('new', 0.016681695220919748), ('great', 0.01217312894499
5492), ('pizza.', 0.012173128944995492), ('one', 0.011722272317403066), ('yor
k', 0.01127141568981064), ('good', 0.01127141568981064), ('place', 0.00946798
9179440937), ('delicious', 0.009017132551848512), ('pizza,', 0.00856627592425
6087), ('eat', 0.007664562669071236), ('cheese', 0.007213706041478809), ('rea
lly', 0.007213706041478809), ('pizza!', 0.007213706041478809), ("joe's", 0.00
6762849413886384), ('thin', 0.005861136158701533), ('de', 0.00541027953110910
8), ('pepperoni', 0.005410279531109108), ('got', 0.004959422903516681), ('sty
le', 0.004959422903516681), ('time', 0.004959422903516681), ('like', 0.004508
566275924256), ('get', 0.004508566275924256), ('fresh', 0.004508566275924256)
, ('delicious.', 0.004508566275924256), ('quick', 0.0040577096483318306), ('e
ver', 0.0040577096483318306), ('taste', 0.0040577096483318306), ('crust', 0.0
036068530207394047), ('lot', 0.0036068530207394047), ('even', 0.0036068530207
394047), ('make', 0.0036068530207394047), ('worth', 0.0036068530207394047), (
'super', 0.0036068530207394047), ('try', 0.0031559963931469793), ('slice.', 0
.0031559963931469793), ('white', 0.0031559963931469793), ('piece', 0.00315599
63931469793), ('nice', 0.0031559963931469793), ('perfect', 0.0031559963931469
793), ('supreme', 0.002705139765554554), ('also', 0.002705139765554554), ('cr
ispy', 0.002705139765554554), ('average', 0.002705139765554554), ('tasty', 0.
002705139765554554), ('say', 0.002705139765554554), ('la', 0.0027051397655545
54), ('love', 0.002254283137962128)]
```

top50bigram

The code defines a function named `bigram` that calculates and prints the top 50 bigrams by frequency in a given list of words. Here's a summary of what the code does:

1. The function takes a single argument `words`, which is the list of words for which we want to calculate the top 50 bigrams.
2. The `BigramCollocationFinder.from_words()` function is used to create a bigram collocation finder from the words. This generates all possible bigrams from the list of words.
3. The `ngram_fd.items()` method is called on the bigram finder to retrieve the bigram frequencies as a list of (bigram, frequency) pairs.
4. The bigram frequencies are sorted in descending order based on the frequency using the `sorted()` function and a lambda function as the key.
5. The sorted list of bigram-frequency pairs is sliced to retrieve the top 50 pairs using the expression `sorted_bigrams[:50]`.
6. The function then iterates over the top 50 bigrams and their frequencies, printing them in the format: "word1 word2: frequency".

To use this function, you can pass a list of words as an argument, and it will calculate and print the top 50 bigrams by frequency.

```
In [66]:  from nltk.collocations import BigramCollocationFinder
```

```
In [77]:  def bigram(words):
              # Generate bigrams
              finder = BigramCollocationFinder.from_words(words)

              # Calculate bigram frequencies
              bigram_frequencies = finder.ngram_fd.items()

              # Sort bigrams by frequency in descending order
              sorted_bigrams = sorted(bigram_frequencies, key=lambda x: x[1], reverse=

              # Extract the top 50 bigrams
              top_50_bigrams = sorted_bigrams[:50]

              # Print the top 50 bigrams and their frequencies
              for bigram, frequency in top_50_bigrams:
                  print(f"{bigram[0]} {bigram[1]}: {frequency}")
```

```python
In [83]:  print("Bigrams for Doc1")
          print()
          top50_bigrams_doc1 = bigram(finalwords_doc1)
          print()
          print("Bigrams for Doc2")
          print()
          top50_bigrams_doc2 = bigram(finalwords_doc2)
```

```
Bigrams for Doc1

best pizza: 30
new york: 25
joe's pizza: 8
one best: 8
york pizza: 7
pizza new: 6
slice pizza: 6
york style: 6
great pizza.: 5
new york.: 5
pizza good: 5
great pizza: 5
pizza place: 5
pizza slice: 5
like pizza: 4
piece pizza: 4
thin crust: 4
cheese pizza: 4
time square: 4
pizza really: 4
great pizza,: 4
style pizza: 4
delicious pizza: 4
good pizza,: 4
pizza pizza: 4
time square.: 3
good place: 3
style pizza.: 3
pizza thin: 3
hot pizza: 3
pizza spot: 3
good pizza: 3
new york,: 3
white pizza: 3
pizza ever: 3
york slice: 3
pizza best: 3
place eat: 3
place small: 3
pizza delicious.: 3
```

```
institution time: 3
new york!: 3
hammer pizza: 2
pizza und: 2
pizza taste: 2
pizza like: 2
want eat: 2
one piece: 2
pizza slice.: 2
near time: 2

Bigrams for Doc2

best pizza: 31
new york: 25
one best: 8
joe's pizza: 8
york pizza: 6
pizza new: 6
pizza good: 6
york style: 6
great pizza.: 5
slice pizza: 5
pizza really: 5
new york.: 5
pizza slice: 5
pizza place: 5
great pizza: 4
like pizza: 4
good pizza,: 4
great pizza,: 4
cheese pizza: 4
delicious pizza: 4
style pizza: 4
thin crust: 4
piece pizza: 4
time square: 4
pizza spot: 3
white pizza: 3
good pizza: 3
pizza delicious.: 3
style pizza.: 3
place eat: 3
pizza. delicious: 3
hot pizza: 3
good place: 3
place small: 3
pizza thin: 3
york slice: 3
pizza best: 3
```

```
institution time: 3
new york!: 3
new york,: 3
time square.: 3
york city: 2
better pizza: 2
pizza delicious: 2
quick service: 2
best slice: 2
slice ever: 2
pizza. slice: 2
slice whole: 2
whole pizza: 2
```

top 50 bigrams by their Mutual Information scores

The provided code defines a function named `bigram_withMI` that calculates and prints the top 50 bigrams by their Mutual Information (MI) scores in a given list of words. Here's a summary of what the code does:

1. The function takes a single argument `words`, which is the list of words for which we want to calculate the top 50 bigrams.
2. The `BigramCollocationFinder.from_words()` function is used to create a bigram collocation finder from the words. This generates all possible bigrams from the list of words.
3. The `score_ngrams()` method is called on the bigram finder with the `BigramAssocMeasures.mi_like` argument to calculate the Mutual Information (MI) scores for the bigrams.
4. The `finder.ngram_fd.items()` generator is used to retrieve the bigram-frequency pairs from the finder.
5. The `filtered_bigrams` list comprehension filters the bigrams based on their frequency, keeping only those with a frequency of 5 or higher.
6. The bigram scores are sorted in descending order based on the score using the `sorted()` function and a lambda function as the key.
7. The sorted list of bigram-score pairs is sliced to retrieve the top 50 pairs using the expression `sorted_bigrams[:50]`.
8. The function then iterates over the top 50 bigrams and their scores, printing them in the format: "word1 word2: score (rounded to 4 decimal places)".

To use this function, you can pass a list of words as an argument, and it will calculate and print the top 50 bigrams by their Mutual Information (MI) scores.

In [78]:
```python
from nltk.metrics import BigramAssocMeasures
def bigram_withMI(words):
    # Generate bigrams
    finder = BigramCollocationFinder.from_words(words)

    # Calculate Mutual Information (MI) scores
    bigram_scores = finder.score_ngrams(BigramAssocMeasures.mi_like)

    # Filter bigrams by frequency
    filtered_bigrams = [bigram for bigram, freq in finder.ngram_fd.items() i

    # Sort bigrams by their Mutual Information scores in descending order
    sorted_bigrams = sorted(bigram_scores, key=lambda x: x[1], reverse=True)

    # Extract the top 50 bigrams
    top_50_bigrams = sorted_bigrams[:50]

    # Print the top 50 bigrams and their Mutual Information scores
    for bigram, score in top_50_bigrams:
        print(f"{bigram[0]} {bigram[1]}: {score:.4f}")
```

In [98]:
```python
print("Bigrams with MI for Doc1")
print()
top50_bigrams_with_MI_doc1 = bigram_withMI(finalwords_doc1)
print()
print("Bigrams with MI Doc2")
print()
top50_bigrams_with_MI_doc2 = bigram_withMI(finalwords_doc2)
```

```
Bigrams with MI for Doc1

new york: 16.8919
best pizza: 2.8632
time square: 1.4545
15-minute queue.: 1.0000
2dollari cadauna): 1.0000
2dollari/cadauna) fate: 1.0000
abstatten. diese: 1.0000
across country.: 1.0000
alike, stand: 1.0000
alive. personally,: 1.0000
alla grandissima!: 1.0000
allright juice: 1.0000
already planning: 1.0000
amazing!! alot: 1.0000
americans. enjoyed: 1.0000
amount saltiness,: 1.0000
argentina knew: 1.0000
arrive hungry: 1.0000
```

```
attend fast,: 1.0000
baked charcoal: 1.0000
balance cheese,: 1.0000
bar none.: 1.0000
barely fit: 1.0000
base. miss: 1.0000
based tastes,: 1.0000
be. obvious: 1.0000
became model: 1.0000
beeline tradition.: 1.0000
beer taken: 1.0000
besuch abstatten.: 1.0000
bezeichnet werden.: 1.0000
birre ghiacciate: 1.0000
bite heaven: 1.0000
block away: 1.0000
bottle water,: 1.0000
bryant park.: 1.0000
buensima mereci: 1.0000
can't believe: 1.0000
caprese, white,: 1.0000
care much.: 1.0000
carrie bradshaw: 1.0000
cena alla: 1.0000
chain originally: 1.0000
charcoal oven,: 1.0000
cheesey! margarita: 1.0000
cher du: 1.0000
chilli fake: 1.0000
cities. usually: 1.0000
closer dark: 1.0000
coda che: 1.0000


Bigrams with MI Doc2

new york: 16.8919
best pizza: 3.1747
time square: 1.4545
15-minute queue.: 1.0000
2dollari cadauna): 1.0000
2dollari/cadauna) fate: 1.0000
abstatten. diese: 1.0000
across country.: 1.0000
alike, stand: 1.0000
alive. personally,: 1.0000
alla grandissima!: 1.0000
allright juice: 1.0000
already planning: 1.0000
amazing!! alot: 1.0000
amount saltiness,: 1.0000
```

```
argentina knew: 1.0000
arrive hungry: 1.0000
attend fast,: 1.0000
baked charcoal: 1.0000
balance cheese,: 1.0000
bar none.: 1.0000
barely fit: 1.0000
base. miss: 1.0000
based tastes,: 1.0000
became model: 1.0000
beeline tradition.: 1.0000
beer taken: 1.0000
besuch abstatten.: 1.0000
bezeichnet werden.: 1.0000
birre ghiacciate: 1.0000
bite heaven: 1.0000
block away: 1.0000
bottle water,: 1.0000
bryant park.: 1.0000
buensima mereci: 1.0000
busy. buena: 1.0000
can't believe: 1.0000
caprese, white,: 1.0000
care much.: 1.0000
carrie bradshaw: 1.0000
cena alla: 1.0000
chain originally: 1.0000
charcoal oven,: 1.0000
cheesey! margarita: 1.0000
cher du: 1.0000
chilli fake: 1.0000
cities. usually: 1.0000
closer dark: 1.0000
coda che: 1.0000
cold, reheat: 1.0000
```

2b) Are there any problems with the word or bigram lists that you found? Could you get a better list of bigrams?

The code calculates the top 50 bigrams by frequency and the top 50 bigrams by their Mutual Information (MI) scores. These are good approaches to identify important bigrams in the data. However, I could have experimented with different measures for scoring bigrams, such as pointwise mutual information (PMI) or chi-square, to see if they yield better results for my specific analysis

2c) How are the top 50 bigrams by frequency different from the top 50 bigrams scored by Mutual Information?

The top 50 bigrams by frequency and the top 50 bigrams scored by mutual information are different. Here's why:

1. Top 50 bigrams by frequency: This refers to the 50 most frequent pairs of consecutive words in the text. These bigrams are determined based on their raw occurrence count. So, the bigrams with the highest count will be considered the most frequent. The frequency-based approach focuses on the sheer occurrence of bigrams, regardless of their significance or contextual relevance.

2. Top 50 bigrams scored by Mutual Information: Mutual information is a statistical measure that quantifies the association between two random variables. In this case, mutual information is used to assess the association between words in a bigram. The mutual information score captures the degree of dependence between the two words in a bigram, taking into account the frequency of their individual occurrences and their joint occurrence. It considers the occurrence of the bigram in relation to the expected occurrence based on the individual word frequencies.

The difference between the top 50 bigrams by frequency and those scored by mutual information lies in the criteria used to rank them. While frequency-based ranking simply looks at the raw counts, mutual information takes into account the statistical dependency between the words in a bigram. As a result, bigrams with higher mutual information scores may not necessarily be the most frequent, but they are likely to exhibit a stronger association or contextual significance compared to the frequency-based approach.

To summarize, the top 50 bigrams by frequency prioritize raw occurrence counts, while the top 50 bigrams scored by mutual information consider the statistical association and relevance of the bigrams. Therefore, the two sets of bigrams are likely to differ in their composition.

d) If you modify the stop word list, or expand the methods of filtering, describe that here.

The provided code defines a function filter_stopwords that filters words from a given list based on several conditions as given below (I used alpha filter and stop words filter along with symbols)

3 Problem: Comparing the style or characteristics between two authors or works based on the difference in the top 50 bigrams by frequency and the top 50 bigrams scored by mutual information.

To analyze the texts and answer the problem, let's assume Doc1 and Doc2 represent two different literary works by different authors.

1. Comparison of Top 50 Bigrams by Frequency:

   - We can observe that both Doc1 and Doc2 share common bigrams such as "best pizza" and "new york," indicating a similarity in the topics discussed.
   - The occurrence of bigrams like "joe's pizza," "york style," "slice pizza," "cheese pizza," and "thin crust" suggests a focus on specific pizza-related details in both documents.
   - Doc1 has bigrams like "time square," "great pizza," and "delicious pizza" with relatively higher frequencies, indicating an emphasis on the experience and quality of pizza in a particular location.
   - Doc2, on the other hand, has bigrams like "pizza good," "pizza really," and "pizza place" with higher frequencies, suggesting a focus on overall pizza quality and places to eat.

2. Comparison of Top 50 Bigrams scored by Mutual Information:

   - Both Doc1 and Doc2 have the same bigrams scored by mutual information, indicating a shared relevance and association between those bigrams and the overall text.
   - The bigrams "new york" and "best pizza" have the highest mutual information scores, implying that these combinations of words are highly informative and

characteristic of the texts.

- Additional bigrams like "time square" and "15-minute queue" also have relatively high mutual information scores, suggesting they contribute to the distinctiveness of the texts.
- The presence of unique bigrams, such as "15-minute queue," "2dollari cadauna)," and "abstatten. diese," indicates specific linguistic features or contextual information in each document.

Based on this analysis, we can infer the following regarding the style or characteristics of the two literary works:

- Both documents share common themes related to pizza and New York.
- Doc1 emphasizes the experience and quality of pizza, particularly in Time Square, with references to great and delicious pizza.
- Doc2 focuses on the overall goodness of pizza and mentions specific places to eat.
- The mutual information scores provide insight into distinctive bigrams that contribute to the unique style and content of each document, potentially indicating differences in tone, language use, or specific details.

Overall, this analysis allows us to compare the texts and identify similarities and differences in their style or characteristics based on the top 50 bigrams by frequency and mutual information.

In [ ]: