

DSC 540-T302 Data Preparation

Week 1&2

Question 4.b.

***** Task 1 *****

Create a string variable which can hold multiline text. Paste the text you just copied into it.

str_text = """It is a truth universally acknowledged, that a single man in possession of a good fortune, must be in want of a wife.

However little known the feelings or views of such a man may be on his first entering a neighbourhood, this truth is so well fixed in the minds of the surrounding families, that he is considered as the rightful property of some one or other of their daughters.

"My dear Mr. Bennet," said his lady to him one day, "have you heard that Netherfield Park is let at last?"

Mr. Bennet replied that he had not.

"But it is," returned she; "for Mrs. Long has just been here, and she told me all about it."

Mr. Bennet made no answer.

"Do not you want to know who has taken it?" cried his wife impatiently.

"You want to tell me, and I have no objection to hearing it."

This was invitation enough.

"Why, my dear, you must know, Mrs. Long says that Netherfield is taken by a young man of large fortune from the north of England; that he came down on Monday in a chaise and four to see the place, and was so much delighted with it that he agreed with Mr. Morris immediately; that he is to take possession before Michaelmas, and some of his servants are to be in the house by the end of next week."

"What is his name?"

"Bingley."

"Is he married or single?"

"Oh! single, my dear, to be sure! A single man of large fortune; four or five thousand a year. What a fine thing for our girls!"

"How so? how can it affect them?"

"My dear Mr. Bennet," replied his wife, "how can you be so tiresome! You must know that I am thinking of his marrying one of them."

"Is that his design in settling here?"

"Design! nonsense, how can you talk so! But it is very likely that he may fall in love with one of them, and therefore you must visit him as soon as he comes."

"I see no occasion for that. You and the girls may go, or you may send them by themselves, which perhaps will be still better, for as you are as handsome as any of them, Mr. Bingley might like you the best of the party."

"My dear, you flatter me. I certainly have had my share of beauty, but I do not pretend to be anything extraordinary now. When a woman has five grown-up daughters, she ought to give over thinking of her own beauty."

"In such cases, a woman has not often much beauty to think of."

"But, my dear, you must indeed go and see Mr. Bingley when he comes into the neighbourhood."

"It is more than I engage for, I assure you."

"But consider your daughters. Only think what an establishment it would be for one of them. Sir William and Lady Lucas are determined to go, merely on that account, for in general, you know, they visit no newcomers. Indeed you must go, for it will be impossible for us to visit him, if you do not."

"You are over scrupulous, surely. I dare say Mr. Bingley will be very glad to see you; and I will send a few lines by you to assure him of my hearty consent to his marrying whichever he chooses of the girls; though I must throw in a good word for my little Lizzy."

"I desire you will do no such thing. Lizzy is not a bit better than the others; and I am sure she is not half so handsome as Jane, nor half so good-humoured as Lydia. But you are always giving her the preference."

"They have none of them much to recommend them," replied he; "they are all silly and ignorant like other girls; but Lizzy has something more of quickness than her sisters."

"Mr. Bennet, how can you abuse your own children in such a way? You take delight in vexing me. You have no compassion on my poor nerves."

"You mistake me, my dear. I have a high respect for your nerves. They are my old friends. I have heard you mention them with consideration these twenty years at least."

"Ah, you do not know what I suffer."

"But I hope you will get over it, and live to see many young men of four thousand a year come into the neighbourhood."

"It will be no use to us, if twenty such should come, since you will not visit them."

"Depend upon it, my dear, that when there are twenty, I will visit them all."

Mr. Bennet was so odd a mixture of quick parts, sarcastic humour, reserve, and caprice, that the experience of three-and-twenty years had been insufficient to make his wife understand his character. Her mind was less difficult to develop. She was a woman of mean understanding, little information, and uncertain temper. When she was discontented, she fancied herself nervous. The business of her life was to get her daughters married; its solace was visiting and news."""

```
type(str_text)
```

```
len(str_text)
```

[1]:

```
4473
```

[2]:

```
# ***** Task 2 *****
```

```
# Now use string methods to get rid of all the newlines (\n or \r)
```

```
# also remove all the symbols including !,?,.,;,, etc.
```

```
new_str_text = str_text.strip()
```

```
clean_str_text = ""
```

```
import re
```

```
clean_str_text = re.sub(r'[?|$|.!|"|,|;|:]',r'',new_str_text)
```

```
print(clean_str_text)
```

It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife

However little known the feelings or views of such a man may be on his first entering a neighbourhood this truth is so well fixed in the minds of the surrounding families that he is considered as the rightful property of some one or other of their daughters

“My dear Mr Bennet” said his lady to him one day “have you heard that Netherfield Park is let at last”

Mr Bennet replied that he had not

“But it is” returned she “for Mrs Long has just been here and she told me all about it”

Mr Bennet made no answer

“Do not you want to know who has taken it” cried his wife impatiently

“You want to tell me and I have no objection to hearing it”

This was invitation enough

“Why my dear you must know Mrs Long says that Netherfield is taken by a young man of large fortune from the north of England that he came down on Monday in a chaise and four to see the place and was so much delighted with it that he agreed with Mr Morris immediately that he is to take possession before Michaelmas and some of his servants are to be in the house by the end of next week”

“What is his name”

“Bingley”

“Is he married or single”

“Oh single my dear to be sure A single man of large fortune four or five thousand a year What a fine thing for our girls”

“How so how can it affect them”

“My dear Mr Bennet” replied his wife “how can you be so tiresome You must know that I am thinking of his marrying one of them”

“Is that his design in settling here”

“Design nonsense how can you talk so But it is very likely that he may fall in love with one of them and therefore you must visit him as soon as he comes”

"I see no occasion for that You and the girls may go or you may send them by themselves which perhaps will be still better for as you are as handsome as any of them Mr Bingley might like you the best of the party"

"My dear you flatter me I certainly have had my share of beauty but I do not pretend to be anything extraordinary now When a woman has five grown-up daughters she ought to give over thinking of her own beauty"

"In such cases a woman has not often much beauty to think of"

"But my dear you must indeed go and see Mr Bingley when he comes into the neighbourhood"

"It is more than I engage for I assure you"

"But consider your daughters Only think what an establishment it would be for one of them Sir William and Lady Lucas are determined to go merely on that account for in general you know they visit no newcomers Indeed you must go for it will be impossible for us to visit him if you do not"

"You are over scrupulous surely I dare say Mr Bingley will be very glad to see you and I will send a few lines by you to assure him of my hearty consent to his marrying whichever he chooses of the girls though I must throw in a good word for my little Lizzy"

"I desire you will do no such thing Lizzy is not a bit better than the others and I am sure she is not half so handsome as Jane nor half so good-humoured as Lydia But you are always giving her the preference"

"They have none of them much to recommend them" replied he "they are all silly and ignorant like other girls but Lizzy has something more of quickness than her sisters"

"Mr Bennet how can you abuse your own children in such a way You take delight in vexing me You have no compassion on my poor nerves"

"You mistake me my dear I have a high respect for your nerves They are my old friends I have heard you mention them with consideration these twenty years at least"

"Ah you do not know what I suffer"

"But I hope you will get over it and live to see many young men of four thousand a year come into the neighbourhood"

"It will be no use to us if twenty such should come since you will not visit them"

"Depend upon it my dear that when there are twenty I will visit them all"

Mr Bennet was so odd a mixture of quick parts sarcastic humour reserve and caprice that the experience of three-and-twenty years had been insufficient to make his wife understand his character Her mind was less difficult to develop She was a woman of mean understanding little information and uncertain temper When she was discontented she fancied herself nervous The business of her life was to get her daughters married its solace was visiting and news

[3]:

***** Task 3 *****

```
word_list = clean_str_text.split()  
len(word_list)
```

[3]:

[4]:

```
# **** Task 4 ****
```

```
for word in word_list:  
    if word_list.count(word) > 1:  
        word_list.remove(word)  
#len(word_list)  
  
unique_word_list = set(word_list)  
#unique_word_list  
len(unique_word_list)
```

367

[4]:

```
# **** Task 5 ****  
# Can you now count how many times each of the unique words appeared in the cleaned text?  
# What is the most efficient way to do it?
```

```
dic = {}  
unique_word_dic = dict.fromkeys(word_list)  
for word in word_list:  
    if unique_word_dic[word] is None:  
        unique_word_dic[word] = 1  
    else:  
        unique_word_dic[word] += 1  
print(unique_word_dic)
```

```
{'It': 1, 'universally': 1, 'acknowledged': 1, 'possession': 2, 'However': 1, 'known': 1, 'feelings': 1, 'views': 1, 'ma  
n': 3, 'first': 1, 'entering': 1, 'neighbourhood': 1, 'this': 1, 'truth': 1, 'well': 1, 'fixed': 1, 'minds': 1, 'surrounding':  
1, 'families': 1, 'considered': 1, 'rightful': 1, 'property': 1, 'some': 2, 'their': 1, 'said': 1, 'lady': 1, 'day': 1, "have":  
1, 'Netherfield': 2, 'Park': 1, 'let': 1, 'last": 1, 'had': 3, 'is": 1, 'returned': 1, "for": 1, 'just': 1, 'here': 1, 'told': 1, 'all  
': 2, 'about': 1, 'made': 1, 'answer': 1, "Do": 1, 'want': 2, 'who': 1, 'taken': 2, 'cried': 1, 'impatiently': 1, 'tell': 1, 'o  
bjection': 1, 'hearing': 1, 'it": 1, 'This': 1, 'invitation': 1, 'enough': 1, "Why": 1, 'Mrs': 1, 'Long': 1, 'says': 1, 'fortu  
ne': 2, 'from': 1, 'north': 1, 'England': 1, 'came': 1, 'down': 1, 'Monday': 1, 'chaise': 1, 'place': 1, 'delighted': 1, 'ag  
reed': 1, 'Morris': 1, 'immediately': 1, 'he': 7, 'before': 1, 'Michaelmas': 1, 'servants': 1, 'house': 1, 'the': 10, 'end':  
1, 'next': 1, 'week": 1, "What": 1, 'name": 1, "Bingley": 1, 'or': 3, 'single": 1, "Oh": 1, 'sure': 2, 'A': 1, 'single': 1,  
'large': 1, 'four': 2, 'What': 1, 'a': 11, 'fine': 1, 'our': 1, 'girls": 1, "How": 1, 'affect': 1, "My": 2, 'Mr': 5, 'Bennet": 1,  
'his': 6, 'wife': 2, "how": 1, 'tiresome': 1, 'must': 5, "Is": 1, 'design': 1, 'settling': 1, 'here": 1, "Design": 1, 'nonsen  
se': 1, 'how': 2, 'talk': 1, 'so': 4, 'likely': 1, 'that': 5, 'fall': 1, 'love': 1, 'therefore': 1, 'soon': 1, 'comes": 1, 'see': 4,  
'occasion': 1, 'may': 1, 'themselves': 1, 'which': 1, 'perhaps': 1, 'still': 1, 'any': 1, 'them': 5, 'might': 1, 'best': 1, 'pa  
rty": 1, 'flatter': 1, 'certainly': 1, 'share': 1, 'beauty': 2, 'pretend': 1, 'be': 5, 'anything': 1, 'extraordinary': 1, 'now  
': 1, 'five': 1, 'grown-up': 1, 'ought': 1, 'give': 1, 'thinking': 1, 'beauty": 1, "In": 1, 'cases': 1, 'often': 1, 'of": 1, 'dea  
r': 3, 'indeed': 1, 'Bingley': 2, 'comes': 1, 'is': 3, 'engage': 1, 'you": 1, 'consider': 1, 'daughters': 2, 'Only': 1, 'think  
': 1, 'an': 1, 'establishment': 1, 'would': 1, 'one': 1, 'Sir': 1, 'William': 1, 'Lady': 1, 'Lucas': 1, 'determined': 1, 'go':  
2, 'merely': 1, 'account': 1, 'general': 1, 'know': 2, 'they': 1, 'no': 4, 'newcomers': 1, 'Indeed': 1, 'for': 4, 'impossib  
le': 1, 'us': 2, 'him': 2, 'not": 1, 'scrupulous': 1, 'surely': 1, 'dare': 1, 'say': 1, 'will': 7, 'very': 1, 'glad': 1, 'you': 9, 's  
end': 1, 'few': 1, 'lines': 1, 'by': 1, 'assure': 1, 'hearty': 1, 'consent': 1, 'to': 7, 'marrying': 1, 'whichever': 1, 'choos  
es': 1, 'of': 8, 'though': 1, 'I': 7, 'throw': 1, 'good': 1, 'word': 1, 'my': 5, 'Lizzy": 1, "I": 1, 'desire': 1, 'do': 2, 'thing':  
1, 'bit': 1, 'better': 1, 'others': 1, 'am': 1, 'handsome': 1, 'as': 2, 'Jane': 1, 'nor': 1, 'half': 1, 'good-humoured': 1, 'Ly  
dia': 1, 'But': 1, 'always': 1, 'giving': 1, 'preference": 1, "They": 1, 'none': 1, 'much': 1, 'recommend': 1, 'replied':  
1, "they": 1, 'silly': 1, 'and': 5, 'ignorant': 1, 'like': 1, 'other': 1, 'girls': 1, 'but': 1, 'Lizzy': 1, 'has': 1, 'something': 1,  
'more': 1, 'quickness': 1, 'than': 1, 'sisters": 1, "Mr": 1, 'Bennet': 2, 'can': 1, 'abuse': 1, 'own': 1, 'children': 1, 'in':  
2, 'such': 2, 'way': 1, 'take': 1, 'delight': 1, 'vexing': 1, 'You': 1, 'have': 3, 'compassion': 1, 'on': 1, 'poor': 1, 'nerves  
": 1, "You": 1, 'mistake': 1, 'me': 1, 'high': 1, 'respect': 1, 'your': 1, 'nerves': 1, 'They': 1, 'old': 1, 'friends': 1, 'hear  
d': 1, 'mention': 1, 'with': 1, 'consideration': 1, 'these': 1, 'at': 1, 'least": 1, "Ah": 1, 'not': 2, 'what': 1, 'suffer": 1,
```

[5]:

"But': 1, 'hope': 1, 'over': 1, 'it': 2, 'live': 1, 'many': 1, 'young': 1, 'men': 1, 'thousand': 1, 'year': 1, 'into': 1, 'neigh
bourhood": 1, "It': 1, 'use': 1, 'if': 1, 'twenty': 2, 'should': 1, 'come': 1, 'since': 1, 'visit': 2, 'them": 1, "Depend': 1
, 'upon': 1, 'when': 1, 'there': 1, 'are': 1, 'all": 1, 'odd': 1, 'mixture': 1, 'quick': 1, 'parts': 1, 'sarcastic': 1, 'humour'
: 1, 'reserve': 1, 'caprice': 1, 'experience': 1, 'three-and-twenty': 1, 'years': 1, 'been': 1, 'insufficient': 1, 'make': 1
, 'understand': 1, 'character': 1, 'Her': 1, 'mind': 1, 'less': 1, 'difficult': 1, 'develop': 1, 'She': 1, 'woman': 1, 'mean'
: 1, 'understanding': 1, 'little': 1, 'information': 1, 'uncertain': 1, 'temper': 1, 'When': 1, 'she': 2, 'discontented': 1
, 'fancied': 1, 'herself': 1, 'nervous': 1, 'The': 1, 'business': 1, 'her': 2, 'life': 1, 'was': 2, 'get': 1, 'married': 1, 'its': 1,
'solace': 1, 'visiting': 1, 'news': 1}

[6]:

```
top_words = sorted(unique_word_dic.items(), key=lambda key_val_tuple: key_val_tuple[1], reverse=True)
top_words[:25]
```

[6]:

```
[('a', 11),
 ('the', 10),
 ('you', 9),
 ('of', 8),
 ('he', 7),
 ('will', 7),
 ('to', 7),
 ('I', 7),
 ('his', 6),
 ('Mr', 5),
 ('must', 5),
 ('that', 5),
 ('them', 5),
 ('be', 5),
 ('my', 5),
 ('and', 5),
 ('so', 4),
 ('see', 4),
 ('no', 4),
 ('for', 4),
 ('man', 3),
 ('had', 3),
 ('or', 3),
 ('dear', 3),
 ('is', 3)]
```