

Project: File Handling & String Manipulation

Basic checks

```
In [3]: #import pandas library
import pandas as pd

In [13]: file = open("dataset/conv.txt", "r")
print (file.read())

WILL: Iâ€™ve never seen wildlings do a thing like this. Iâ€™ve never seen a thing like this, not ever in my life.

WAYMAR ROYCE: How close did you get?

WILL: Close as any man would.

GARED: We should head back to the wall.

ROYCE: Do the dead frighten you?

GARED: Our orders were to track the wildlings. We tracked them. They wonâ€™t trouble us no more.

ROYCE: You donâ€™t think heâ€™ll ask us how they died? Get back on your horse.

WILL: Whatever did it to them could do it to us. They even killed the children.

ROYCE: Itâ€™s a good thing weâ€™re not children. You want to run away south, run away. Of course, they will behead you as a deserter â€¦ If I donâ€™t catch you
u first, get back on your horse. I wonâ€™t say it again.

ROYCE: Your dead men seem to have moved camp.

WILL: They were here.

GARED: See where they went.

ROYCE: What is it?

JON: Go on. Fatherâ€™s watching.

JON: And your mother.

SEPTA MORDANE: Fine work, as always. Well done.

SANSA: Thank you.

SEPTA MORDANE: I love the detail that youâ€™ve managed to get in this corners.

NED: And which one of you was a marksman at ten? Keep practicing, Bran. Go on.

JON: Donâ€™t think too much, Bran.

ROBB: Relax your bow arm.

CASSEL: Lord Stark. My lady. A guardsman just rode in from the hills. Theyâ€™ve captured a deserter from the Nightâ€™s Watch.

NED: Get the lads to saddle their horses.

CATELYN: Do you have to?

NED: He swore an oath, Cat.

CASSEL: The law is law, my lady.

NED: Tell Bran heâ€™s coming, too.

CATELYN: Ned. Ten is too young to see such things.

NED: He wonâ€™t be a boy forever. And winter is coming.

ROBB: Lad, go run back and get the rest.

NED: King of the Andals and the First Men â€¦

JON: Father will know if you do.

NED: Lord of the Seven Kingdoms and protector of the realm, I, Eddard of the House Stark, Lord of Winterfell and Warden of the North, sentence you to die.

JON: You did well.

NED: You understand why I did it?

BRAN: Jon said he was a deserter.

NED: But do you understand why I had to kill him?

BRAN: Our way is the old way?

NED: The man who passes the sentence should swing the sword.

BRAN: Is it true he saw the White Walkers?

NED: The White Walkers have been gone for thousands of years.

BRAN: So he was lying?

NED: A madman sees what he sees.

JON: What is it?

THEON: Mountain lion?

NED: There are no mountain lions in these woods.

THEON: Itâ€™s a freak.

NED: Itâ€™s a direwolf.

NED: Tough old beast.

ROBB: There are no direwolves south of the Wall.

JON: Now there are five.

JON: You want to hold it?

BRAN: Where will they go? Their motherâ€™s dead.

CASSEL: They donâ€™t belong down here.

NED: Better a quick death. They wonâ€™t last without their mother.

THEON: Right. Give it here.

BRAN: No!

ROBB: Put away your blade.

THEON: I take orders from your father, not you.

BRAN: Please, father!

NED: Iâ€™m sorry, Bran.

JON: Lord Stark? There are five pups. One for each of the Stark children. The direwolf is the sigil of your House. They were meant to have them.

NED: You will train them yourselves. You will feed them yourselves. And if they die, you will bury them yourselves.

BRAN: What about you?

JON: Iâ€™m not a Stark. Get on.

ROBB: What is it?

CERSEI: What if Jon Arryn told someone?

JAIME: But who would he tell?

CERSEI: My husband.

JAIME: If he told the king, both our heads would be skewered on the city gates by now. Whatever Jon Arryn knew or didnâ€™t know, it died with him. And Robert
will choose a new Hand of the king, someone to do his job while heâ€™s off fucking boars and hunting whores. Or is it the other way around? And life will go o
n.

CERSEI: You should be the Hand of the king.

JAIME: Thatâ€™s an honor I can do without. Their days are too long, their lives are too short.

CATELYN: All these years and I still feel like an outsider when I come here.

NED: You have five northern children. Youâ€™re not an outsider.

CATELYN: I wonder if the old gods agree.

NED: Itâ€™s your gods with all the rules.

CATELYN: I am so sorry, my love.

NED: Tell me.

ROBB: I hear the prince is a right royal prick.

THEON: Think of all those southern girls he gets to stab with his right royal prick.

ROBB: Go on, Tommy, shave him good. Heâ€™s never met a girl he likes better than his own hair.

CATELYN: Gods, but they grow fast.

BRAN: I saw the king! Heâ€™s got hundreds of people!

CATELYN: How many times have I told you No climbing!

BRAN: But heâ€™s coming right now! Down our road!

CATELYN: I want you to promise me. No more climbing.

NED: Your Grace.

ROBERT: Youâ€™ve got fat.

ROBERT: Cat!

CATELYN: Your Grace.

ROBERT: Nine years. Why havenâ€™t I seen you? Where the hell have you been?

NED: Guarding the North for you, Your Grace. Winterfell is yours.

ARYA: Whereâ€™s the Imp?

SANSA: Will you shut up?

ROBERT: Who have we here? You must be Robb.

ARYA: Arya.

ROBERT: Ooh. Show us your muscles. Youâ€™ll be a soldier.

ARYA: Thatâ€™s Jaime Lannister. The queenâ€™s twin brother.

SANSA: Would you please shut up.

NED: My queen.

CATELYN: My queen.

ROBERT: Take me to your crypt. I want to pay my respects.

CERSEI: Weâ€™ve been riding for a month, my love. Surely the dead can wait.

ROBERT: Ned.

ARYA: Whereâ€™s the Imp?

CERSEI: Where is our brother? Go find the little beast.

NED: Tell me about Jon Arryn.

ROBERT: One minute he was fine, and then â€¦ Burned right through him, whatever it was. I loved that man.

NED: We both did.

ROBERT: He never had to teach you much, but me â€¦ You remember me at 16? All I wanted to do was crack skulls and fuck girls. He showed me what was what.

NED: Aye.
```

```
In [14]: file = pd.read_csv("dataset/conv.txt", delimiter = ';', header=None, index_col=None)

In [15]: file

Out[15]:
```

	0	1
0	WILL	I've never seen wildlings do a thing like thi...
1	WAYMAR ROYCE	How close did you get?
2	WILL	Close as any man would.
3	GARED	We should head back to the wall.
4	ROYCE	Do the dead frighten you?
...
106	NED	Tell me about Jon Arryn.
107	ROBERT	One minute he was fine, and then ... Burned rig...
108	NED	We both did.
109	ROBERT	He never had to teach you much, but me ... You...
110	NED	Aye.

111 rows x 2 columns

```
In [16]: #replace the column names
file.columns=['Speaker', 'Dialogue']

In [17]: file.head()

Out[17]:
```

	Speaker	Dialogue
0	WILL	I've never seen wildlings do a thing like thi...
1	WAYMAR ROYCE	How close did you get?
2	WILL	Close as any man would.
3	GARED	We should head back to the wall.
4	ROYCE	Do the dead frighten you?

```
In [18]: #Unique names of the speaker
file['Speaker'].unique()

Out[18]: array(['WILL', 'WAYMAR ROYCE', 'GARED', 'ROYCE', 'JON', 'SEPTA MORDANE',
'SANS', 'NED', 'ROBB', 'CASSEL', 'CATELYN', 'BRAN', 'THEON',
'ERSEI', 'JAIME', 'ROBERT', 'ARYA'], dtype=object)

In [19]: #No of dialogues for each speakers
file['Speaker'].value_counts()

Out[19]:
```

NED	27
CATELYN	19
JON	18
BRAN	18
ROBERT	9
ROBB	7
ROYCE	5
CERSEI	5
THEON	5
WILL	4
ARYA	4
CASSEL	3
SANSA	3
JAIME	3
GARED	3
SEPTA MORDANE	2
WAYMAR ROYCE	1
Name: Speaker, dtype: int64	

```
In [20]: #Count the unique number of the speakers
speaker_set = set(file['Speaker'])
count=0
for speaker in speaker_set:
count = count+1
print("Total unique speakers are:",count)

Total unique speakers are: 17

In [ ]: #Speakers list
#speaker_list = list(file['Speaker'])
#speaker_list

In [25]: #Dialogues list
#dialogue_list = list(file['Dialogue'])
#dialogue_list

In [22]: #Dialogues Test
#cleaning the unnecessary words and make them in lower case

dialogue_list_lower=[]
dialogue_list_final=[]

for lower_d in dialogue_list:
str_lower=lower_d.lower()
dialogue_list_lower.append(str_lower)

dialogue_list_lower_clean = [re.sub("[->()\"#@;%;<>{}'+-!?,-]", "", _) for _ in dialogue_list_lower]

for d in dialogue_list_lower_clean:
str_test(' '.join(dict.fromkeys(d.split()))
dialogue_list_final.append(str_test)

dialogue_list_final

Out[22]: ['I've never seen wildlings do a thing like this not ever in my life',
'how close did you get',
'close as any man would',
'we should head back to the wall',
'do the dead frighten you',
'our orders were to track the wildlings we tracked them they won't trouble us no more',
'you don't think he'll ask us how they died get back on your horse',
'whatever did it to them could do us they even killed the children',
'it's a good thing we're not children you want to run away south of course they will behead as deserter if i don't catch first get back on your horse won't s
ay it again',
'your dead men seem to have moved camp',
'they were here',
'see where they went',
'what is it',
'go on father's watching',
'and your mother',
'fine work as always well done',
'thank you',
'i love the detail that you've managed to get in this corners',
'and which one of you was a marksman at ten keep practicing bran go on',
'don't think too much bran',
'relax your bow arm',
'lord stark my lady a guardsman just rode in from the hills they've captured deserter night's watch',
'get the lads to saddle their horses',
'do you have to',
'he swore an oath cat',
'the law is my lady',
'tell bran he's coming too',
'ned ten is too young to see such things',
'he won't be a boy forever and winter is coming',
'lad go run back and get the rest',
'king of the andals and first men',
'father will know if you do',
'lord of the seven kingdoms and protector realm i eddard house stark winterfell warden north sentence you to die',
'you did well',
'you understand why i did it',
'jon said he was a deserter',
'but do you understand why i had to kill him',
'our way is the old',
'the man who passes sentence should swing sword',
'is it true he saw the white walkers',
'the white walkers have been gone for thousands of years',
'so he was lying',
'a madman sees what he',
'what is it',
'mountain lion',
'there are no mountain lions in these woods',
'it's a freak',
'it's a direwolf',
'tough old beast',
'there are no direwolves south of the wall',
'now there are five',
'you want to hold it',
'where will they go their mother's dead',
'they don't belong down here',
'better a quick death they won't last without their mother',
'right give it here',
'no',
'put away your blade',
'i take orders from your father not you',
'please father',
'i'm sorry bran',
'lord stark there are five pups one for each of the children direwolf is sigil your house they were meant to have them',
'you will train them yourselves feed and if they die bury',
'what about you',
'i'm not a stark get on',
'what is it',
'what if jon arryn told someone',
'but who would he tell',
'my husband',
'i hear the king both our heads would be skewered on city gates by now whatever jon arryn knew or didn't know it died with him and robert will choose a n
ew hand of someone to do his job while he's off fucking boars hunting whores is other way around life go',
'you should be the hand of king',
'i love the detail that you've managed to get in this corners',
'that's an honor i can do without their days are too long lives short',
'all these years and i still feel like an outsider when come here',
'you have five northern children you're not an outsider',
'i wonder if the old gods agree',
'it's your gods with all the rules',
'i am so sorry my love',
'tell me',
'i hear the prince is a right royal prick',
'think of all those southern girls he gets to stab with his right royal prick',
'go on tommy shave him good he's never met a girl he likes better than his own hair',
'gods but they grow fast',
'i saw the king he's got hundreds of people',
'how many times have i told you no climbing',
'but he's coming right now down our road',
'i want you to promise me no more climbing',
'your grace',
'you've got fat',
'cat',
'your grace',
'nine years why haven't i seen you where the hell have been',
'guarding the north for you your grace winterfell is yours',
'where's the imp',
'will you shut up',
'who have we here you must be robb',
'arya',
'ooh show us your muscles you'll be a soldier',
'that's jaime lannister the queen's twin brother',
'would you please shut up',
'my queen',
'my queen',
'take me to your crypt i want pay my respects',
'we've been riding for a month my love surely the dead can wait',
'ned',
'where's the imp',
'where is our brother go find the little beast',
'tell me about jon arryn',
'one minute he was fine and then burned right through him whatever it i loved that man',
'we both did',
'he never had to teach you much but me remember at 16 all i wanted do was crack skulls and fuck girls showed what',
'aye']
```

Task: A sample script of Game of Thrones was taken from the page and stored in the file conv.txt in the dataset provided. Your task is to read the file and do the following:

1. Find out the number of unique characters in the sample conversation?

```
In [30]: #listing of speakers
#in conv.txt file, we observe that the dialogues and characters are separated by a '\'.
# on which we will use split function and append the separated characters in a blank list.

characters_in_csv= []

# read all the lines_in csv in txt file
text_file = open("dataset/conv.txt","r", encoding='utf-8');
complete_text = text_file.readlines()
text_file.close()

# ignore all lines with only newline character and choose the rest
for line in complete_text:
    if (line == '\n'):
        pass
    # list of characters in text file obtained by splitting with delimiter ": "
    else:
        character=line.split(":")[0] #until
        # For unique values, skip the name if it is already in the list
        if character in characters_in_csv :
            pass
        else :
            characters_in_csv.append(character)

In [31]: print("Number of unique characters in the script : ", len(characters_in_csv))
print("The Unique Names of the speakers : ")

for name in characters_in_csv :
    print(name, end = ', ')

Number of unique characters in the script : 17
The Unique Names of the speakers :
WILL, WAYMAR ROYCE, GARED, ROYCE, JON, SEPTA MORDANE, SANSA, NED, ROBB, CASSEL, CATELYN, BRAN, THEON, CERSEI, JAIME, ROBERT, ARYA,
```

2. Create a new text file for each of the characters with their name and store the unique words said by them in their respective file. Store one word in one line.

```
In [32]: # In the given text file we will again split the character and append them in a blank list, which will create
# a nested list comprising list of each dialogue along with
# its speaker on the 1st and 8th index respectively.

name_set = set(characters_in_csv)

conversation = []

with open('dataset/conv.txt', 'r',encoding='utf-8') as text:
    for i in text:
        if i[0]!='\n':
            pass
        else:
            conversation.append(i.split(':')[1])

In [36]: #Dialogues of speakers
print("Select the name of the speaker with in the list")
array = []
with open ('dataset/conv.txt',encoding='utf-8')as f:
    [array.append(line) for line in f.readlines()]
name_of_the_speakers = input()
for a in array:
    if (Name_of_the_speakers in a): #for full conversations remove the line
        print(a)

Select the name of the speaker with in the list
WILL
WILL: I've never seen wildlings do a thing like this. I've never seen a thing like this, not ever in my life.

WILL: Close as any man would.

WILL: Whatever did it to them could do it to us. They even killed the children.

WILL: They were here.
```

```
In [34]: # So,Now we will compare the 8th index containing names in "conversation" with the names in "name_set" created in above question
# in a nested for loop. If the 8th index will match, 1st index of that list will be converted to string and
# appended in the blank string "dialogue_str" in each loop.

for i in name_set:
    dialogue_str = ""
    for j in conversation:
        if j[0] == i:
            dialogue_str = dialogue_str+j[1:]

# Now "dialogue_str" will be cleaned and all the words will be converted to uppercase/ lowercase in order to avoid
# repetition due to case difference and then splitted to individual words.

    dialogue_str = dialogue_str.replace(' ','').replace(".", "").replace(",", "").replace("?", "").upper()
    dialogue_split = dialogue_str.split()

# Convert these individual words present inside "dialogue_split" in to a set for getting unique words named "dialogue_set".
    dialogue_set = set(dialogue_split)

# Insert these words from "dialogue_set" using for loop, creating respective text files by creating a file writing path
# in "w" mode, then close the file for the "for loop".
    file = open(i+'.txt','w')
    for k in dialogue_set:
        file.write(k + '\n')
    file.close()

In [ ] :
```