```
""" it.304 week 10 Project - 7 pillars of python review
Created on Sat Oct 22 07:42:57 2022 - Updated 10/31/22
@author: 17574 => b.hogan@snhu.edu """

#=> 7 Pillars of Python Project - it.304 System Analysis and Design
#=> Good writing is good thinking, and good programming is mad-hatting
#=> due date =  when you complete the exercises to the best of your ability

"""Over the past few weeks have you have taken the time to learn Python data
structures, data pack/unpack, data transformation, conditionals, iterators,
functions, and building your own object-oriented classes. Whew!

This is a significant undertakening and all autobots.it304 have performed
amazingly well. You have shown your capacity to think different and
"sticktoitness." You are doing the work others find challenging.

Now let's take Shakespeare from the bottom up. In this challenge you will
work directly with text files and duplicate the spreadsheet tables from
weeks 5-8. You will import, transform, iterate, use conditionasl, and end
with a simple object asking user what they want to read and giving it to them.
Use your skills and make it so. ~brianh

p.s. Please provide quality and thoughtful answers to all questions. I
encourage you to write as much as you want, ask me questions back, and use
this as your opportunity to build a piece of evidence you can put in your
portfolio and showcase to an employer. I am happy to help you expand items if
you would like to do.

#=> EXAM Part 0 - fill in the blanks

     obj_Name   | charcter code      |   explicit code
      --------- | ------------------ | -----------------------------
  i) mytuple =  |  (, )              |=> mytuple = tuple(myobject)
 ii) mylist =   |                    |=> mylist  =
iii) mydict =   |                    |=> mydict  =
 iv) myset  =   |                    |=> myset   =
  v) dataframe =|                    |=> df      =
 vi) mystring = |                    |=> mystring=
"""


# data files here in either a zip or individual files. if you download git
# desktop you can clone the folder and it installs the file in your documents.
# sorry we did not have time to review this common feature.
# folder address
#https://github.com/bbe2/IT.304.Fall.2022/tree/Shakespeare-Corpus
# zip file in folder address
#https://github.com/bbe2/IT.304.Fall.2022/blob/Shakespeare-
Corpus/shakespeare_txt_fullname.zip
```

```
#---------------------------------------------------------------------------
#=> EXAM Part I    - replicate shakespeare spreadsheet from wks 5-8
#---------------------------------------------------------------------------
  #    Situation: unfortunately your business customer cant read the data b/c
  #    they dont know how to open text files in WORD. PLEASE help them out!
  #      1a) get the names of the play
  #      1b) create a list of play script data
  #      1c) count total words in script and titles
  #      1d) create numeric indexID for each: hint -> list(range(999))
  #      1f) DARN-it! there is not play type information. What should u do?
  #      1e) create a dictionary that matchs weeks 5-8 input spreadsheet
  #          => title, script, type, id
  #      1g) send dict to df, df to spreadsheet, email to me

#---------------------------------------------------------------------------
#=> EXAM Part II   - create summary report by play type
#---------------------------------------------------------------------------
  #       Total all script words and title words by 3 play types
  #       send to df to spreadsheet and email to me

#---------------------------------------------------------------------------
#=> EXAM Part III  - ask user what play they want to read and email the data
#---------------------------------------------------------------------------
#       3a) Create an object with one or two functions.
#           Ask user what play they want to read.
#           Figure out a minimum of 1 other useful piece of information
#           to display or include in user report.
#           Have function export data and send me data file.

#==========================================================================
#=> EXAM Part I    - replicate shakesprease spreadsheet from wks 5-8
#==========================================================================
#==========================================================================

import pandas as pd                               #dataframe library
#df = pd.read_excel("."),df.to_dict(),pd.DataFrame.from_dict(mydict)
import numpy as np                                #num library
import matplotlib.pyplot as plt                   #visualization library
import os                                         #op system, #msft.os=\\
import sys                                        # sys.exit()
os.chdir('C:\\Users\\17574\\Desktop\\data_it304\\shakespeare_txt_name')
path = 'C:\\Users\\17574\\Desktop\\data_it304\\shakespeare_txt_name'
                            # QUESTION: how is a path different from directory?
                            # urANSWER:
                            #
```

```
#-------------------------
#=> I.a GET NAMES OF PLAYS
#-----------
mylist_filenames = os.listdir(path)
                                # QUESTION: hey! what os.listdir?
                                # What function tells u quickly ? (hint->SOS)
                                # copy/paste 1st 2 lines of that info here.
                                # urANSWER:
                                #
print(mylist_filenames)
print(type(mylist_filenames))
len(mylist_filenames)
                                # this is a Jackson inspired question...
                                #QUESTION: why 37 ? Anything missing?
                                # research on an academic website MIT
                                #urANSWER:
                                #
                                #
                                # As a junior analyst why am i asking you what
                                # is missing? How would it impact decisions
                                # made in the user object in part III.
                                #
                                #
                                #
                                #
#=> this is very useful method to get names and file paths in any directory
#=> you might want to make note of for future project plan building
mylist_playnames= []
for file in os.listdir(path):
    #print(path+ "\\" + file)
    #next = path + "\\" + file
    filename = file.split(".")
    justname = filename[0]
    mylist_playnames.append(justname)
print(len(mylist_playnames))
                                # QUESTION: what is filename[0] doing above?
                                # We didn't talk about this but bagels if
                                # 3 out of 4 get it right!
                                # urANSWER:
                                #
                                #
                                # QUESTION: why might I have you print len()
                                # again when you just did so in prior code ?
                                # urANSWER:
                                #
                                #
print(path+ "\\" + file)
                                # QUESTION: hmmm, you have path name too?
                                # Curious, what could you do with this?
                                # urANSWER:
                                #
                                #
```

```
del file; del filename; del justname
                                # QUESTION: what is important about this code?
                                # urANSWER:
                                #
                                #
print(type(mylist_playnames))
print(type(mylist_playnames[0]))
print(mylist_playnames)
                                # QUESTION: please explain how data is packed
                                # urANSWER:
                                #
                                #
                                #
                                # QUESTION: u need to count words or characters
                                # so how would prefer data to be packed?
                                # urANSWER:

#-------------------------------------
#=> I.b CREATE A LIST OF SCRIPT DATA
#=> we did not do 'with file open' in class. I am helping you out >:)
#------------------------
mylist_data= []                 # => megasaurus
for file in os.listdir(path):
    with open(file,'r') as data:    # 'r' is the parameter for read info
        mylist_data.append(data.readlines()) #.readlines() is a method
print(mylist_data)
len(mylist_data)
                                # QUESTION: explain what code is doing
                                # urANSWER:
                                #
                                #
                                # QUESTION: show me code for
                                # help data.readlines method information.
                                # copy/paste 1st 2 lines of that data here too.
#=> YOUR CODE HERE <-
                                # urANSWER:
                                #
                                #
```

```
#----------------------------------------------------------------------
#=> I.c GET TOTAL Title WORDS
#   Use an iterator
#   Use string method .split() to count words vs characters like class code
#   ex: len(mylist_playnames[0].split())
#---------------------------------------
print(type(mylist_playnames))
print(type(mylist_playnames[0]))
                          # QUESTION: how data packed and can you count words?
                          # urANSWER:
                          #
print(mylist_playnames[0])
len(mylist_playnames[0].split())
                          # QUESTION: did you get 4?
                          # if not your files were stored different then mine
                          # but you will get same totals for reports
                          # my 1st play = A Midsummer Nights Dream, 4 words
                          # urANSWER:
                          #
                          #

#=> YOUR CODE HERE <----get title total words
#=> build a list of total words per title; hint you only need 1 for loop
#=> write a print statement for min, max, and total words in titles
'''#=> 'ANSWER: min, max, total = 1, 7, 121'''
mylist_words_title = []
#=> YOUR CODE HERE
```

```
#----------------------------------------------------------------------
#=> I.c GET TOTAL script total  WORDS
print(type(mylist_data))
print(type(mylist_data[0]))
                                # QUESTION: how is data packed? And,
                                # explain how data is packed differently in
                                # a list than a string.
                                # urANSWER:
                                #
                                #
print(type(str(mylist_data[0])))
                                # QUESTION: explain what str() function does
                                # urANSWER:
                                #
                                #
   #hint= https://docs.python.org/3/library/functions.html#built-in-functions
print(len(mylist_data[0].split()))
                               # QUESTION: why is this not working?
                                # What did you learn about data pack/unpack
                                # in this exercise?
                                # urANSWER:

print(len(str(mylist_data[0]).split()))
#16026
                                # QUESTION: tell me what python object .split()
                                # is a method of, where you got the
                                # information, and how long it took you.
                                # urANSWER:
                                #
                                #
#------------------------------------
#=> I.c GET TOTAL script
#=> hint: use a for loop inside a for loop;
    # exterior loops each play in a list
        # interior loops the words in the play
            # for i in mylist:
            #    for word in i:
            #        <more code here to add to new list>
#----------------------

#=> build a list of total words per title; hint you need 2 loops
#=> write a print statement for min, max, and total words in script
'''#=> ANSWER: min, max, total = 14495 71154 878202 '''

mylist_words_script= []
#=> YOUR CODE HERE <- print and put information in a comment here





print(min(mylist_words_script),max(mylist_words_script),sum(mylist_words_script))
```

```
#-------------------------------------------------------------------
#=> I.d create numeric indexID for each play: hint -> list(range(999))
# #print and put information in a comment here
#------------------------------

mylist_id = []
#=> YOUR CODE HERE <-
#copy and paste output here too as a comment



print(type(mylist_id))
print(mylist_id)
#-------------------------------------------------------------------
# 1f) DARN-it! there is no play type information. How solve this?
# ok - your own here, but a minion built you a couple lists
# hint: use for loop and conditionals
#------------------------
comedy = ['A Midsummer Nights Dream', 'Alls Well That Ends Well',
'As You Like It',
 'Comedy of Errors', 'Cymbeline', 'Hamlet',
'Loves Labours Lost',
 'Measure for Measure', 'Much Ado About Nothing',
 'Pericles',
 'The Merchant of Venice', 'The Merry Wives of Windsor',
 'The Taming of the Shrew', 'The Tempest',
 'Twelfth Night', 'Two Gentlemen of Verona ', 'Winters Tale']
history = [ 'Henry IV part 1', 'Henry IV part 2',
 'Henry V', 'Henry VI part 1', 'Henry VI part 2', 'Henry VI part 3',
 'Henry VIII',  'Richard II',
 'Richard III',  'The Life and Death of King John']
tragedy = ['Antony and Cleopatra', 'The Tragedy of Coriolanus', 'Macbeth',
 'Timon of Athens', 'Titus Andronicus', 'Troilus and Cressida',
  'Othello the Moore of Venice','Romeo and Juliet',
 'The Life and Death of Julius Caesar','King Lear',]
print(len(comedy), len(history),len(tragedy))
#17 10 10


mylist_play_type = []
#=> YOUR CODE HERE <-
#=> hint - you dont have to do this manually if use: if i in mylist_playnames



print(mylist_play_type)
print(len(mylist_play_type))
```

```
#------------------------------------------------------------------
# 1f) create a dictionary that matchs weeks 5-8 input spreadsheet
  #          => title, script, type, id, words_script, words_title
#-------------------------------------

mydict = {}
#=> YOUR CODE HERE <-




print(mydict)
#------------------------------------------------------------------
#=> 1g) send dict to df, export to spreadsheet, email to me
#------------------

df1 = pd.DataFrame()
#=> YOUR CODE HERE <-





#==========================================================================
#=> EXAM Part II   - create summary report by play type
#==========================================================================
#==========================================================================
  #       Total all script words and title words by 3 play types
  #       send to df to spreadsheet and email to me
  #     Answer:
          # comedy  + history   + tragedy
          # 371235  + 236820    + 270147    = 878202
          # 53      + 35        + 33        = 121

mydict2 = {}
#=> YOUR CODE HERE <-




print(mydict2)
```

```
#---------------------------------------------------------------------------
#=> EXAM Part III  - ask user what play they want to read and email the data
#---------------------------------------------------------------------------
#       3a) Create an object with one or two functions.
#           Ask user what play they want to read.
#           Figure out a minimum of 1 other useful piece of information
#           to display or include in user report.
#           Have function export data
#           you can simply email me the file!
```

```
#---------------------------------------------------------------------------
#=> EXAM Part IV - sorry your not done yet !
#---------------------------------------------------------------------------

#congrats for making it here! <insert yippe dance>

#commander lamda is very concerned about your ability to manipulate data

#please create
# a tuple with the play names
# a tuple with the play's id
# use the input() function an ask user what play they want to read
# the user only wants to read Moby Dick but all other books must be there
# take the play data they want to read and put it into a list
# Ask user if they want to read the "wrath of kahn?"
# from the list print them the wrath of kahn words
# i will give you a hint in class if you are having trouble
# your amazing !
```