CSCI 182 Homework 1
Due Date: Oct 1st, 2021

All assignments MUST have your name, student ID, course name/number at the beginning of your documents.
Your homework MUST be submitted via Camino with file format and name convention as follows:
HW#_Name_code.zip (for coding part)
If you have any questions, please don't hesitate to contact me :)

You have 8 tasks to finish, please write all codes in a single jupyter-notebook file called "hw1.ipynb" instead of 8 files :)

• **Task 1**: **Define Object**
Define a function called 'task1' to assign items= [1, 2, 3, 4, 5] as a list object, and print the list.

• **Task 2**: **File Reading**
Create a file called 'task2.data.' and type string '1 2 3 4 5 6 7 8 9 10' in it without quotation marks, and then write a function to read the file and load the string as two list-objects items1 = [1, 2, 3, 4, 5], items2 = [6, 7, 8, 9, 10]. Print the two lists (items1 and items2). Read more about .data files here:
https://www.askpython.com/python/examples/read-data-files-in-python

• **Task 3**: **Data Structure**
It is important to be familiar with the functions of a dictionary: items(), keys(), values(), Write a program to print the dictionary using these functions.
Notes: Dictionary has two ways to initialize
data = dict()
data['school'] = 'UAlbany'
data['address'] = '1400 Washington Ave, Albany, NY 12222'
data['phone'] = '(518) 442-3300'

data = {'school': 'UAlbany', 'address': '1400 Washington Ave, Albany, NY 12222', 'phone': '(518) 442-3300'}
Print the results as follows by accessing the defined dictionary:
school: UAlbany
address: 1400 Washington Ave, Albany, NY 12222
phone: (518) 442-3300

• **Task 4**: **Data Serialization**
Use json to store the above dictionary in a file and then load the same and print item, key and values. Read more about json format here:
https://www.w3schools.com/js/js_json_intro.asp
This task tells how to store a dictionary object to a file and then load the dictionary object from the file (Hint: dictionary → json string→ string in a file → json string → dictionary). In python, objects of any type can be mostly saved in json format to a txt file.

You need to be familiar with the following two json functions : json.dumps(object), which dumps an object to a json format (string), json.loads(a json format string), which loads a json format string back to the original object. You also need to store and load these strings from the file. We do not care if the original object is a list, dictionary or any other. json.dumps() can automatically recognize the type.

Note, json.loads(object) can only load an object, not a file.

• **Task 5**: **Data Serialization**
In this task we store a number of different types of objects (e.g., list, dictionary, array) to a file and then load the objects from the file.
Write a function to dump list object items = [1,2,3,4,5] and above dictionary to a file called 'task5.data', and then load them from the same file and print.

• **Task 6**: **Data Preprocessing**
Read the tweets from the file "CrimeReport.txt" and print the id for each tweet.
Here are some functions that you will use in the task: open().readlines(), tweet = json.loads(), print tweet.keys(),
Once you know the keys of the tweet dictionary object, then you can find which key relates to tweet id, and you can then retrieve the id of this specific tweet.

• **Task 7**: **Data Preprocessing: tweets filtering**

INPUT: "CrimeReport.txt"
OUTPUT: a file "task7.data" that stores the 10 most recent tweets
Suggestions:
tweet['created − at'] gives the created time of this tweet. Rank tweets based on the time from the earliest to the most recent. Then we can identify the 10 most recent tweets.
Some example lines that are not directly runnable: [image next page]

```
import datetime
tweets = []
for line in open().readlines():
    tweet = json.loads(line)
    tweets.append(tweet)
    #datetime.datetime.strptime(item['created-at'], '%a %b %d %H:%M:%S +0000 %Y')
    #converts the string format of a date time to the datetime object
sorted_tweets = sorted(tweets, key = lambda item:
    datetime.datetime.strptime(item['created-at'], '%a %b %d %H:%M:%S +0000 %Y'))
# sorted tweets based on time.
f = open('output.txt', 'w')
for tweet in sorted_tweets[-5:]:
    f.write(json.dumps(tweet) + '\')
f.close()
```

Note, when you use the code above, please be careful with the proper indentation and quotation mark. [Hint: You might need to change some values here and there :)]

• **Task 8: File operations**
INPUT :CrimeReport.txt: in this file, each line is a raw tweet json format.
output-folder: where new results will be stored

REQUIREMENT: read tweets and separate these tweets into groups based on the specific hours (Mon-Day-Year-Hour). The tweets related to a specific hour will be stored in a separate file in the folder ”task8-output” with the file name ”Mon-Day-Year-Hour.txt”

OUTPUT: new files generated and stored in the folder ”task8-output”, in which each file stores the tweets corresponding to a specific hour.