Programming for Data Analytics
Numpy

Question 1. Numerical Analysis Exercises using NumPy Bike Dataset

For each of the following questions you will use the bike rental dataset called bike.csv. Where possible use NumPy to answer the questions below.
The following are the details of the various fields in this dataset.

1. instant: record index
2. season : season (1:springer, 2:summer, 3:fall, 4:winter)
3. yr : year (0: 2011, 1:2012)
4. mnth : month (1 to 12)
5. hr : hour (0 to 23)
6. holiday : weather day is holiday or not (extracted from [Web Link])
7. weekday : day of the week
8. workingday : if day is neither weekend nor holiday is 1, otherwise is 0.
9. + weathersit :
    i. 1: Clear, Few clouds, Partly cloudy, Partly cloudy
    ii. 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
    iii. 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
    iv. 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
10. temp : Normalized temperature in Celsius. The values are divided to 41 (max)
11. atemp: Normalized feeling temperature in Celsius. The values are divided to 50 (max)
12. hum: Normalized humidity. The values are divided to 100 (max)
13. windspeed: Normalized wind speed. The values are divided to 67 (max)
14. casual: count of casual users
15. registered: count of registered users
16. cnt: count of total rental bikes including both casual and registered

(i)

Calculate the average temperature value (index 9) for the entire dataset. Note the temperature values in this column have been normalized by dividing by 41.

(ii)

Print out the average number of rental users for all days classified as holidays as well as the average for all days classified as non-holidays. (Note holidays =1 and non-holidays = 0). Holidays attribute is stored at index 5.

(iii)

Write NumPy code that will print out the total number of casual users for each month of the year. You would expect to see an increase in the number of casual users over the summer months and a decline for the winter months.

(iv)

We will now look at the relationship between temperature and the number of rental users. Your code should work out the average number of rental users for the following temperature ranges.

- • 1, 5
- • 6, 10
- • 11, 15
- • 16, 20
- • 21, 25
- • 26, 30
- • 31, 35
- • 36, 40

Remember the temperature values specified in the file have been normalised by dividing by 41.

```
For temp in range  1 to 5 the mean number of casual users was  49.2954545455
For temp in range  6 to 10 the mean number of casual users was  73.6670630202
For temp in range  11 to 15 the mean number of casual users was  130.681770652
For temp in range  16 to 20 the mean number of casual users was  169.066772655
For temp in range  21 to 25 the mean number of casual users was  211.700074516
For temp in range  26 to 30 the mean number of casual users was  242.172678691
For temp in range  31 to 35 the mean number of casual users was  337.473005641
For temp in range  36 to 40 the mean number of casual users was  314.991111111
```

Solution:

```
# -*- coding: utf-8 -*-

"""

Solutions for NumPy  Exercises


"""
```

Question 1

```
import numpy as np



def bicycleDataSolutions():

    data = np.genfromtxt('bike.csv', delimiter=',')

    # Q2 (i) Get average tempature
    allAverages = np.mean(data, axis=0)
    print ("Average Temperature is ", allAverages[9]*41)

    # Q2 (ii)
    print ("Mean number of non-holiday users {}".format(cacluateMeanUsersHoliday(data, 0)))
    print ("Mean number of holiday users {}".format(cacluateMeanUsersHoliday(data, 1)))

    # Q2 (iii) Mean number of casual users per month
    calculateUsersPerMonth(data)

    for temp in range(1, 40, 5):
        analyseTemp(data, temp, temp+4)

def cacluateMeanUsersHoliday(data, dayType):
```

```
    dayTypeSubset = data[data[:,5]==dayType]

    numUsers = dayTypeSubset[:, 15]

    return (np.mean(numUsers))



def calculateUsersPerMonth(data):

    monthName = ["Jan", "Feb", "March", "April", "May", "June", "July", "Aug", "Sept", "Oct",
"Nov", "Dec"]


    for currentMonth in range(1,13):

        booleanRowsForMonth = (data[:, 3] == currentMonth)


        dataForMonth = data[booleanRowsForMonth]


        print ("Total users for month {} is {}".format(monthName[currentMonth-1],
np.sum(dataForMonth[:,13])))




def analyseTemp(data, minValue, maxValue):

    # the temperature values stored in the array are multiplied by 41


    higherTempCondition = (data[:,9]*41)>=minValue

    lowerTempCondition = (data[:,9]*41)<=maxValue


    subset = data[higherTempCondition & lowerTempCondition]
```

```
    meanValue = np.mean(subset[:, 15])

    print ("For temp in range ", minValue, "to", maxValue, "the mean number of casual users was
", meanValue)
```