

AML-2103 Visualization for AI and ML

Final

(20% of Final grade)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Instructor:** | | | Vahid Hadavi, PhD, P.Eng | | |
| **Class:** | | | **AML-2103** | | |
| Please follow Exam Instruction :   * This Exam includes 22 multiple choice questions, and **3** short answer questions. * Read all questions carefully. * Just pencil is acceptable for Scantron sheet. * Write your name and student ID in front page correctly. * Time allowed: 90 **min** | | | | | |
| **Date:** | | …………………………………….…………………….. |
| **Student Name:** | | …………………………………….…………………….. |
| **Student Number:** | | …………………………………….…………………….. |

# Section 1: Multiple Choice Question Mark: 60

1.What will be output for the following code?

import pandas as pd

s = pd.Series([1,2,3,4,5],index = ['a','b','c','d','e'])

print s['a']

A. 1  
B. 2  
C. 3  
D. 4

2.What will be syntax for pandas dataframe?  
A. pandas.DataFrame( data, index, dtype, copy)  
B. pandas.DataFrame( data, index, rows, dtype, copy)  
C. pandas\_DataFrame( data, index, columns, dtype, copy)  
D. pandas.DataFrame( data, index, columns, dtype, copy)

3.Axis 1, in panel represent?  
A. minor\_axis  
B. major\_axis  
C. items  
D. None of the above

4.Which of the following thing can be data in Pandas?  
A. a python dict  
B. an ndarray  
C. a scalar value  
D. All of the above

5.The \_\_\_\_\_\_\_\_ project builds on top of pandas and matplotlib to provide easy plotting of data.  
A. yhat  
B. Seaborn  
C. Vincent  
D. Pychart

6.What will be output for the following code?  
import pandas as pd

import numpy as np

s = pd.Series(np.random.randn(2))

print s.size

A. 0  
B. 1  
C. 2  
D. 3

7.Is anything wrong with the plot below?

# whats_wrong

# A.Nope, everything looks fine!

# B.Yes, the data seems to be messy/wrong

# C.Yes, this data is better represented as a pie chart

# D.Yes, this data is better represented as a vertical bar chart

# 8. What kind of plot(s) would you use to examine the distribution of a numeric variable?

# A.Scatter plot

# B.Boxplot

# C.Bar plot

# D.Histogram

# 9.In what order should you run the following commands in order to generate and display a plot?

# 1.plt.legend()

# 2.x = np.linspace(1, 10, 20)

# 3.plt.show()

# 4.plt.xlabel('X-Label')

# 5.plt.plot(x, label='Example Plot')

# A.4-2-1-3-5

# B.2-5-4-1-3

# C.5-4-1-3-2

# D.1-2-4-3-5

10.Fill in the blanks such that the code block generates the following figure:

**import** matplotlib.pyplot **as** plt

sample\_figure **=** plt**.**figure()

*# 1*

ax1 **=** \_\_\_

*# 2*

ax2 **=** \_\_\_

*# Draw plots*

ax1**.**hist(data)

ax2**.**boxplot(data)

plt**.**show()

# 

A - ax.add\_subplot(122) | 2 - ax.add\_subplot(121)

B - sample\_figure.add\_subplot(121) | 2 - sample\_figure.add\_subplot(122)

C - add\_subplot(121) | 2 - add\_subplot(122)

D - sample\_figure.add(1, 2, 2) | 2 - sample\_figure.add(1, 2, 1)

11.Fill in the blanks such that the code block generates the following figure:

**import** seaborn **as** sns

sns**.**set\_style('darkgrid')

*# 1*

ax **=** \_\_\_(x**=**'cyl', y**=**'mpg', data**=**df)

*# 2*

\_\_\_(xlabel **=** "Cylinders", ylabel**=**'Miles per gallon')

# 

## A. 1 - sns.boxplot | 2 - sns.set

## B. 1 - boxplot | 2 - ax.set

## C. 1 - boxplot | 2 - ax.set\_labels

## D. 1 - sns.boxplot | 2 - ax.set

## 12.[Best way to import the pandas module in your program ?](http://r4r.in/mcqs/mcqs-questions-answers.php?que_id=294&test_id=9)

## A. import pandas

## B. import pandas as p

# C. from pandas import \*

# D. All of the above

## 13.[DataFrame in pandas is](http://r4r.in/mcqs/mcqs-questions-answers.php?que_id=299&test_id=9)

# A. 1 dimensional array

# B. 2 dimensional array

# C. 3 dimensional array

# D. None of the above

# 14.What is the output of the following code?

# print 9//2

# A. 4.5

# B. 4.0

# 4

# Error

# 15. It is possible to convert the Numpy array to list in python ?

# A.Yes

# B.No

# C.Sometimes

# D.None of the above

## 16.[Numpy.array(list), what it does ?](http://r4r.in/mcqs/mcqs-questions-answers.php?que_id=237&test_id=6)

# A.It convert array to list

# B.It convert list to array

# C.It convert array to array

# D.Error

## 17.[Which of the following counts the number of elements in Numpy array ?](http://r4r.in/mcqs/mcqs-questions-answers.php?que_id=242&test_id=6)

# A.count()

# B.return()

# C.shape()

# D.size()

# 18.What is the output for −

# 'you are doing well' [2:999]

# A. 'you are doing well'

# B. ' '

# C. Index error.

# D. 'u are doing well'

# 19.What is output for following code −

# y = [4, 5,1j]

# y.sort()

# A. [1j,4,5]

# B. [5,4,1j]

# C. [4,5,1j]

# D. Type Error

# 20. In the following options which are python libraries which are used for data analysis and scientific computations

# A. Numpy

# B. Scipy

# C. Pandas

# D. All the above

# 21. Which options are correct to create an empty set in Python?

# A. {}

# B. ()

# C. []

# D. set()

# 22. What is output of following code −

# def func(x, ans):

# if(x==0):

# return 0

# else:

# return func(x-1, x+ans)

# print(func(2,0))

# A. 0

# B. 1

# C. 2

# D. 3

# Section 2: Short Answer Questions Mark: 40

# Answer each question in a clear and organized paragraph. Each paragraph should include clear and precise sentences that directly addresses the question prompt. Use complete sentences. The length of your answer should be one to four sentences or, where required, one to two lines. *(LONGER THAN SPECIFIED ANSWERS WILL BE IGNORED)*

# 1.Please Explain what is a bussiness Intelligence dashboard and what are the advantages of a bussiness intelligence dashboard in Python. (10 Marks)

# 2.Name 8 of the steps you may need to apply in data cleaning? (10 marks)

# 3. Starbucks cooperation is an American company founded in 1971 in Seattle. Starbucks has about 182,000 employees across 19,767 company operated and licensed stores in 62 countries. You are hired as a data analytic expert in Starbucks and you want to help Starbucks make better decisions and be able to extract knowledge out of its big data. In less than 10 lines explain your methodology including the data identification, data procurement, data cleaning, data visualization and etc. (20 marks)