

AML-2103 Visualization for AI and ML

Midterm

(20% of Final grade)

**Answers**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Instructor:** | | | Vahid Hadavi, PhD, P.Eng | | |
| **Class:** | | | **AML-2103** | | |
| Please follow Exam Instruction :   * This Exam includes 15multiple choice questions, and **4** 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: 60 **min** | | | | | |
| **Date:** | | …………………………………….…………………….. |
| **Student Name:** | | …………………………………….…………………….. |
| **Student Number:** | | …………………………………….…………………….. |

# Section 1: Multiple Choice Question Mark: 60

# 1. Which is a python package used for 2D graphics?

# A. matplotlib.pyplot

# B. matplotlib.pip

# C. matplotlib.numpy

# D. matplotlib.plt

# Answer : Option A

# 2. Identify the package manager for Python packages, or modules.

# A. Matplotlib

# B. PIP

# C. plt.show()

# D. python package

# Answer : Option B

# 3. Observe the output figure. Identify the coding for obtaining this output.

# 

# A.

# import matplotlib.pyplot as plt

# plt.plot([1,2,3],[4,5,1])

# plt.show()

# B.

# import matplotlib.pyplot as plt

# plt.plot([1,2],[4,5])

# plt.show()

# C.

# import matplotlib.pyplot as plt

# plt.plot([2,3],[5,1])

# plt.show()

# D.

# import matplotlib.pyplot as plt

# plt.plot([1,3],[4,1])

# plt.show()

# Answer : Option A

# 4. Read the code:

# import matplotlib.pyplot as plt

# plt.plot(3,2)

# plt.show()

# Identify the output for the above coding.

# A.

# 

# B.

# 

# C.

# 

# D.

# 

# Answer : Option C

# 5. Identify the right type of chart using the following hints.

# Hint 1: This chart is often used to visualize a trend in data over intervals of time.

# Hint 2: The line in this type of chart is often drawn chronologically.

# A. Line chart

# B. Bar chart

# C. Pie chart

# D. Scatter plot

# Answer : Option A

# 6.Read the statements given below. Identify the right option from the following for pie chart.

# Statement A: To make a pie chart with Matplotlib, we can use the plt.pie() function.

# Statement B: The autopct parameter allows us to display the percentage value using the Python string formatting.

# A. Statement A is correct

# B. Statement B is correct

# C. Both the statements are correct

# D. Both the statements are wrong

# Answer : Option C

# 7. Pandas is an open-source \_\_\_\_\_\_\_ Library? A. Ruby B. Javascript C. Java D. Python

# Ans : D

# 8. Pandas key data structure is called? A. Keyframe B. DataFrame C. Statistics D. Econometrics

# Ans : B

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

# import numpy as np

# s = pd.Series(np.random.randn(4))

# print s.ndim

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

# Ans : B

# 10. 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 mentioned

# Ans :d

# 11. Point out the correct statement. A. If data is a list, if index is passed the values in data corresponding to the labels in the index will be pulled out B. NaN is the standard missing data marker used in pandas C. Series acts very similarly to a array D. None of the mentioned

# Ans : b

# 12.Which of the following input can be accepted by DataFrame? A. Structured ndarray B. Series C. DataFrame D. All of the mentioned

# Ans:D

# 13.If data is an ndarray, index must be the same length as data. A.True B.False

# Answer:a Explanation: If no index is passed, one will be created having values [0, …, len(data) – 1].

# 14. Correct syntax of the reshape() function in Numpy array python is…

# A. array.reshape(shape)

# B. reshape(array,shape)

# C. array.reshape(shape)

# D. reshape(shape,array)

# E. reshape(shape)

# Ans : A

# 15. Which of the following find the maximum number in the Numpy array ?

# A. max(array)

# B. array.max()

# C. max(array)

# D. array(max)

# Ans : A

# 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)*

# Explain data Wrangling? (10 Marks)

# Ans.:

# Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis.

# Data wrangling is the process of gathering, selecting, and transforming data to answer an analytical question. Also known as data cleaning or “munging”, legend has it that this wrangling costs analytics professionals as much as 80% of their time, leaving only 20% for exploration and modeling

# Explain an industrial example where al the different below advanced NumPy operations are done: Filtering, Sorting, Combining and Reshaping.

# (10 Marks)

# Ans: This could be an industrial example like tweeter data, Starbucks data and etc.

# Explain advantages and drawbacks of pandas over NumPy, please (10 marks)

# Pandas is an open-source library built on top of numpy providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language. It allows for fast analysis and data cleaning and preparation. operations on NumPy arrays can be significantly faster than operations on Pandas series. NumPy arrays can be used in place of Pandas series when the additional functionality offered by Pandas series isn't critical. ... Running the operation on NumPy array has achieved another four-fold improvement. Pandas is so fast because it uses numpy under the hood. Numpy implements highly efficient array operations.

# Describe 4 different types of plots and explain the use case of each. (10 Marks)

*Pie Chart:*

* *Display relative proportions of multiple****classes****of data.*
* *size of the circle can be made proportional to the****total****quantity it represents.*
* *summarize a large data set in visual form.*
* *be visually simpler than other types of graphs.*
* *permit a visual check of the reasonableness or****accuracy****of calculations.*

*Density Plot :*

* *A density plot shows the distribution of a numerical variable. It takes only set of numeric values as input. It is really close to a histogram. Since it is a really common dataviz technique, most of the dataviz libraries allow to draw it.*

*Bubble chart Python*

* *A bubble plot is a scatterplot where a third dimension is added: the value of an additional variable is represented through the size of the dots. You need 3 numerical variables as input: one is represented by the X axis, one. by the Y axis, and one by the size.*

*Heatmap*

* *Is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors. The seaborn python package allows the creation of annotated heatmaps which can be tweaked using Matplotlib tools as per the creator's requirement.*