



الجامعة اللبنانية
كلية الإعلام

الجامعة اللبنانية
كلية الإعلام
الفرع الأول

Third Year

Semester VI

Advanced Topics in Data Science

Date:

Duration:

Final Exam

2025 / 2026

Instructor : Dr. K. Danach

SECTION 1: 20 Multiple Choice Questions (MCQs). Each question has only one answer. 1 pt. /question, 20 points in total.

Q1. Which pandas function is used to detect missing values in a DataFrame?

- A. `df.empty()`
- B. `df.isnull()`
- C. `df.checknull()`
- D. `df.missing()`

Q2. Which function is used to remove rows containing missing values?

- A. `df.remove()`
- B. `df.delete_na()`
- C. `df.dropna()`
- D. `df.clean()`

Q3. Which command is used to select a column named "Age" from a DataFrame `df`?

- A. `df(Age)`
- B. `df["Age"]`
- C. `df{Age}`
- D. `df.select("Age")`

Q4. Which function is used to save a DataFrame to a CSV file?

- A. `df.save_csv()`
- B. `df.export_csv()`
- C. `df.to_csv()`

D. `df.write_csv()`

Q5. Which pandas function is used to check for missing values in a DataFrame?

A. `df.empty()`

B. `df.isnan()`

C. `df.isnull()`

D. `df.missing()`

Q6. Which method shows column names, data types, and non-null counts?

A. `df.info()`

B. `df.describe()`

C. `df.details()`

D. `df.check()`

Q7. How are missing values usually represented in pandas?

A. NaN

B. None

C. NA

D. Empty

Q8. Which function is used to detect missing values in a DataFrame?

A. `df.isnull()`

B. `df.isnan()`

C. `df.empty()`

D. `df.checkna()`

Q9. Which function replaces missing values with a specific value?

A. `df.fillna()`

B. `df.replace_na()`

C. `df.setna()`

D. `df.fixna()`

Q10. Which command fills missing values with the column mean?

A. `df.fillna(df.mean())`

B. `df.fillna(mean)`

C. `df.mean().fill()`

D. `df.replace(mean)`

Q11. Which method removes rows containing missing values?

A. `df.remove_na()`

B. `df.dropna()`

C. `df.delete_na()`

D. `df.clean_na()`

Q12. Which pandas function is used to read an Excel file?

A. `pd.read_excel()`

B. `pd.load_excel()`

C. `pd.open_excel()`

D. `pd.import_excel()`

- Q13. Which statistical measure is commonly used to detect outliers?
- A. Median
 - B. Standard Deviation
 - C. Mode
 - D. Mean
- Q14. Which visualization is often used to identify outliers?
- A. Histogram
 - B. Box plot
 - C. Line chart
 - D. Pie chart
- Q15. Which method shows the first rows of a DataFrame by default?
- A. df.head()
 - B. df.sample()
 - C. df.tail()
 - D. df.view()
- Q16. Which command gives a quick look at the dataset content?
- A. df.head()
 - B. df.describe()
 - C. df.info()
 - D. df.clean()
- Q17. Which command is used to remove rows with missing values in pandas?
- A. df.remove()
 - B. df.dropna()
 - C. df.delete()
 - D. df.clearna()
- Q18. Which function is used to display the first rows of a DataFrame?
- A. df.start()
 - B. df.first()
 - C. df.head()
 - D. df.top()
- Q19. Which command returns the column names of a DataFrame?
- A. df.headers
 - B. df.columns
 - C. df.names
 - D. df.fields
- Q20. Which scikit-learn function is used to split data into training and testing sets?
- A. split_data()
 - B. train_test_split()
 - C. data_split()
 - D. test_train()

SECTION 2: 10 True and False questions. 1.5 point/question, 16.5 points in total

- T 1. NumPy is a Python library mainly used for numerical computations and array operations.
- T 2. The np.array() function is used to create an array in NumPy.
- T 3. The mean() function calculates the average value of elements in an array.
- T 4. The shape attribute gives the dimensions of a NumPy array.
- F 5. A NumPy array can contain elements of different data types in the same column without conversion.
- F 6. The zeros() function creates an array filled with zeros.
- T 7. The max() function is used to find the largest value in an array.
- T 8. The reshape() function changes the dimensions of an array without modifying its data.
- F 9. NumPy arrays are generally slower than Python lists for numerical calculations.
- T 10. The arange() function generates a sequence of numbers within a specified range.

SECTION 3: Data Science Under Python, 8 points/task, 64 points in total

Dataset Description

You are given a CSV file named employee_data.csv with the following structure:

Column Name	Description
WorkHours	Number of work hours per week
Experience	Years of professional experience
Projects	Number of completed projects
TrainingHours	Training hours attended per month
Performance	Good or Poor (target variable)

Example of the data:

WorkHours	Experience	Projects	TrainingHours	Performance
-----------	------------	----------	---------------	-------------

40	5	12	6	Good
25	1	3	2	Poor
NaN	4	8	5	Good
45	7	15	7	Good
20	NaN	2	1	Poor

The dataset may contain missing values and outliers.

Question

Using Python, pandas, and scikit-learn, perform the following tasks:

Task 1 — Read and Explore the Data

1. Read the CSV file into a pandas DataFrame.
2. Display the first 5 rows of the dataset.
3. Display the last 5 rows of the dataset.

Task 2 — Dataset Summary

1. Display statistical information about the dataset.
2. Display information about columns, data types, and missing values.

Task 3 — Handling Missing Values

1. Identify columns containing missing values.
2. Remove rows with missing values.
3. Fill missing values using the mean of each numerical column.

Task 4 — Outlier Detection (IQR Method)

1. Choose the column WorkHours.
2. Calculate Q1, Q3, and the Interquartile Range (IQR).

3. Identify outliers using the IQR rule.
4. Remove the detected outliers from the dataset.

Task 5 — Data Splitting

1. Split the dataset into training and testing sets.
2. Use 80% for training and 20% for testing.
3. Set a random state for reproducibility.

Task 6 — Classification Model

1. Create a Logistic Regression or Decision Tree Classifier model.
2. Train the model using the training data.
3. Predict employee performance using the test data.
4. Display the predicted values.

Task 7 — Model Evaluation

1. Generate the Confusion Matrix.

Task 8 — Model Evaluation

1. Calculate the Accuracy Score.
2. Calculate Precision, Recall, and F1-Score.
3. Display all evaluation results clearly.