Python for ML Mid Term

100 Marks Deadline: 26th October, 11:59 PM
90 Marks Deadline: 28th October, 11:59 PM

Submission Instructions:

1. Watch the assignment instruction video carefully before starting.
2. Create a **Google Colaboratory (.ipynb)** file in your Google Drive.
3. **Make Text Cell and give the following information:**
   Your registered email address (the one used for your Phitron account).
4. Write all answers for all questions in **a single file**.
5. For each question, create a **Text cell** with the question number and a **Code cell** with your solution.
6. Print or display all required outputs clearly so that graders can verify results.
7. Share the Colab file with **"Anyone with the link"** and permission set to **"Viewer"**.
8. Submit only the Colab link.

Note: Each Question is carrying 10 marks.

---

0) Starter Code (Run First)

```
# You must run this code and use that data set for all the questions
import numpy as np
import pandas as pd

np.random.seed(42)

ids = np.arange(1, 11)
ages = np.random.randint(18, 60, 10)
salaries = np.random.randint(30000, 90000, 10)
departments = np.array(["HR", "IT", "Finance", "IT", "HR", "Sales", "Finance", "IT", "Sales", "HR"])

DF = pd.DataFrame({
"id": ids,
"age": ages,
"salary": salaries,
"dept": departments
})

DF.to_csv("employees.csv", index=False)
print("Sample Data Created and Saved as employees.csv")
```

---

1) NumPy — 5 Questions

Q1. Create NumPy arrays from the `ages` and `salaries` data generated above. Print both arrays and display their `dtype`, `ndim`, `shape`, and `size`.

Q2. Using the `salaries` array, find and print the highest and lowest salary values. Also, calculate and print the average salary and age using NumPy functions.

Q3. From the `ages` array, filter and print all ages greater than 30 using a boolean condition. Then count and print how many employees are older than 30.

Q4. Create a new NumPy array that increases every employee's age by 5 years (without modifying the original array). Print the new updated ages array.

Q5. Using NumPy, calculate the total salary expense (sum of all salaries) and the difference between the maximum and minimum salary. Print both results.

---

2) Pandas — 5 Questions

Q6. Load the `employees.csv` file into a Pandas DataFrame. Display the first 5 rows, and check basic info using `info()` and summary statistics using `describe()`.

Q7. Display only the `id`, `age`, and `salary` columns. Then show the last 3 rows using `tail()`.

Q8. Filter the DataFrame to show only employees who work in the `IT` department. Print the result and show the total number of IT employees.

Q9. Sort the DataFrame by `salary` in descending order and display the top 3 highest-paid employees along with their department and age.

Q10. Replace all `salary` values greater than 80000 with 80000 using `loc`. Then calculate and print the new average salary of all employees after replacement.