AI 306 Homework 1 (15%)

Deadline: Thursday 06th March

Dr. Mohammed Temraz
Academic year 2024-2025

Homework Files

- ./AI306_HW1.pdf — homework questions (this file).
- ./ Employee Attrition.csv: — data file for the questions.

Expected output files

- ./Homework-01-StudentNumber.ipynb — Python notebook programs.
- ./Homework-01-Report-StudentNumber.docx — Report in word format with answers.

Requirements

- Python 3.8+, pandas 1.3+, numpy 1.20+, sklearn 0.24+.

In this homework, you will be working with "Employee Attrition and Performance" dataset. The dataset consists of several information about different aspects of an employee's profile. The aim is to predict employee attrition (whether an employee leaves the company or not). The target/class variable is "Attrition", Yes or No, where Yes indicates that employee leaves the company and No indicates that employee will not leave the company.

1) Write a Python code to answer the following (Write the answer and the code under each question): (4 points)
   a. Show the dimension of the dataset.
   b. Identify any duplicate record (s). How many?
   c. keeping one duplicated record, delete the other record (s) from the dataset.
   d. What is the dimension of the data frame after removing the duplicates?

2) Write a Python code to answer the following (and show your codes): (5 points)
   a. How many male and female employees are there in the dataset?
   b. What is the mean value of Hourly Rate feature?
   c. What percent of the employees with a job role "Research Scientist"?
   d. What percent of the employees with an education field "Life Sciences"?
   e. What is the mean Monthly Rate of a male employee?

3) Write a Python code to answer the following (and show your codes): (4 points)
   a. What type of variable is "EmployeeNumber"?
   b. What is the median of "TotalWorkingYears"?
   c. Find the variance and standard deviation of "TrainingTimesLastYear"
   d. Find the five numbers summary of "HourlyRate"

4) Calculate the median of grouped data in the table (show all steps) (2 points).

Grades	Frequency (# of students)
0 - 20	4
20 - 40	12
40 - 60	9
60 - 80	20
80 - 100	15