Assignment-2

Pandas

Q.1 Create a dataset as follow in the table.

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Q.2 From the above table that you have created, Locate row 0, 4, 7 and 8 using dataframe.

Q.3 Read a csv file, and display it.

Understanding Attribute Types Using Python

Q.4 Understanding Attribute Types Using Python

- Represent the following:
    - Nominal attributes (e.g., names, departments)
    - Binary attributes (e.g., gender, pass/fail)
    - Ordinal attributes (e.g., low < medium < high)
    - Numeric attributes (continuous and discrete)

Q.5 Navigate the DataFrame and do the following task for the table created in question 1:

1. Select row from index 3 to 7.
2. Select row from index 4 to 8, and column 2 to 4.
3. Select all rows with column index 1 to 3 (include index 3 during selection).

Q.6 Read a csv file and display its first five rows.
Note-: Download dataset from https://www.kaggle.com/datasets/uciml/iris)

Q.7 From the csv file (uploaded in the Q.6) delete row 4, and delete column 3. Display the result.

Q.8 Create a sample dataset (employees.csv) containing information about employees in a company.

Employee_ID	Name	Department	Age	Salary	Years_of_Experience	Joining_Date	Gender	Bonus	Rating
101	Alice	HR	29	50000	4	2020-03-15	Female	5000	4.5
102	Bob	IT	34	70000	8	2017-07-19	Male	7000	4.0
103	Charlie	IT	41	65000	10	2013-06-01	Male	6000	3.8
104	Diana	Marketing	28	55000	3	2021-02-10	Female	4500	4.7
105	Edward	Sales	38	60000	12	2010-11-25	Male	5000	3.5

WAP to determine the following using dataset in Q8:

a) Shape (number of rows and columns) of the DataFrame.
b) Summary of the DataFrame that includes the data types and non-null counts for each column.

c) Generate descriptive statistics.
d) Display the first 5 rows and last 3 rows of the dataset.
e) Calculate the following statistics from the dataset:
    i. The average salary of employees.
    ii. The total bonus paid to all employees.
    iii. The youngest employee's age.
    iv. The highest performance rating.
f) Sort the DataFrame by the Salary column in descending order.
g) Add a new column that categorizes employees based on their performance rating:
    i. Excellent for ratings >= 4.5
    ii. Good for ratings >= 4.0 but < 4.5
    iii. Average for ratings < 4.0
h) Identify missing values in the DataFrame.
i) Rename the Employee_ID column to ID.
j) Find all employees who:
    i. Have more than 5 years of experience.
    ii. Belong to the IT department.
k) Modify the dataset by adding a new column, Tax, which deducts 10% of the Salary.
l) Save the modified DataFrame (with added columns) to a new CSV file.