

# Assignment-3

## UCS420 “Cognitive Computing”

### 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 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.4 Read a csv file and display its first five rows.

Note-: Download dataset from <https://www.kaggle.com/datasets/uciml/iris>

Q.5 From the csv file (uploaded in the Q.4) delete row 4, and delete column 3. Display the result.

Q.6 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 Q6:

- Shape (number of rows and columns) of the DataFrame.
- Summary of the DataFrame that includes the data types and non-null counts for each column.
- Generate descriptive statistics.
- Display the first 5 rows and last 3 rows of the dataset.
- Calculate the following statistics from the dataset:
  - The average salary of employees.
  - The total bonus paid to all employees.
  - The youngest employee's age.
  - The highest performance rating.
- Sort the DataFrame by the Salary column in descending order.
- Add a new column that categorizes employees based on their performance rating:
  - Excellent for ratings  $\geq 4.5$
  - Good for ratings  $\geq 4.0$  but  $< 4.5$
  - Average for ratings  $< 4.0$
- Identify missing values in the DataFrame.
- Rename the Employee\_ID column to ID.
- Find all employees who:
  - Have more than 5 years of experience.
  - Belong to the IT department.
- Modify the dataset by adding a new column, Tax, which deducts 10% of the Salary.
- Save the modified DataFrame (with added columns) to a new CSV file.