

Question # 1:

(2x20=40)

The "Playing Tennis" dataset contains weather-related attributes and a binary target variable indicating whether a game of tennis was played.

Dataset Description

- **Features:**
 - **Outlook:** The weather outlook, which can be 'Sunny', 'Overcast', or 'Rain'.
 - **Temperature:** The temperature on the day, categorized as 'Hot', 'Mild', or 'Cool'.
 - **Humidity:** The humidity level, either 'High' or 'Normal'.
 - **Wind:** The wind conditions, either 'Weak' or 'Strong'.
- **Target Variable:**
 - **PlayTennis:** Indicates whether tennis was played ('Yes') or not ('No') on a given day based on the weather conditions.

Example Data Points

Here are a few example records from the dataset:

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Sunny	Mild	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No

Write the Python code to accomplish the following tasks using the "Playing Tennis" dataset. Assume the dataset is stored in a file named `Playtennis.csv`.

1. Load the "Playing Tennis" dataset into a Pandas DataFrame and display the first five rows.

2. Select and display the rows where 'Temperature' is 'Hot' and 'Humidity' is 'High'.

3. Introduce some missing values in the 'Temperature' column and then fill these missing values with the mode of the column.

4. Group the data by 'Outlook' and calculate the mean of 'Temperature' (after converting to numerical values) for each group.

5. Create a new column 'HumidityBinary' that contains 1 if 'Humidity' is 'High' and 0 otherwise.

6. Use the `apply()` function to create a new column 'WeatherScore' where 'Sunny' adds 2, 'Overcast' adds 3, and 'Rain' adds 1 to the score.

7. Split the DataFrame into two separate DataFrames: one with features and one with the target variable. Merge them back together using `pd.concat()`.

8. Create a pivot table that shows the count of days grouped by 'Outlook' and 'playTennis'

9. Reshape the DataFrame using melt() such that all weather conditions are melted into a single column with corresponding values.

~~Sort~~ the DataFrame by 'Temperature' and then by 'Humidity'.

11. Filter the DataFrame to include only rows where 'Wind' is 'Weak' and 'PlayTennis' is 'Yes'.

↳ Convert the 'Outlook' column to categorical type and display the categories.

15. Export the DataFrame to a CSV file named 'playing_tennis_processed.csv'.

12. Assume each row in the dataset represents a day. Add a 'Date' column starting from '2021-01-01' and create a new column 'DayOfWeek' to show the day of the week.

13. Create a new column 'WeatherDescription' that combines 'Outlook', 'Temperature', and 'Humidity' into a single string.

↳ Convert the 'Outlook' column to categorical type and display the categories.

15. Export the DataFrame to a CSV file named 'playing_tennis_processed.csv'.