

## Pandas Dataframe Practice Test

Dataset: (Run these two lines)

```
import seaborn as sns
```

```
penguins = sns.load_dataset('penguins')
```

Data Attributes Info:

- id**: Unique identifier for each penguin, useful for distinguishing observations (Integer/String).
- species**: Species of the penguin ("Adelie", "Chinstrap", "Gentoo"), often used for classification tasks (Categorical).
- island**: Island in the Palmer Archipelago where the penguin was observed ("Biscoe", "Dream", "Torgersen"), helps in geographical analysis (Categorical).
- bill\_length\_mm**: Length of the penguin's bill in millimeters, important for species differentiation (Continuous, Float).
- bill\_depth\_mm**: Depth of the penguin's bill in millimeters, used alongside bill length for species analysis (Continuous, Float).
- flipper\_length\_mm**: Length of the penguin's flipper in millimeters, helps analyze physical characteristics (Continuous, Float).
- body\_mass\_g**: Body mass of the penguin in grams, used to study size and health (Continuous, Float).
- sex**: Sex of the penguin ("Male", "Female", "NA"), useful for comparing characteristics between genders (Categorical).
- year**: Year of the observation, used to analyze trends over time (Integer).

## Questions:

- I. Check if any column in the dataset contains missing values. If yes, display the column names.
- II. Retrieve rows where penguins have an "island" value of "Torgersen" and only display their "flipper\_length\_mm" and "body\_mass\_g".
- III. Filter penguins from the "Dream" island that are male. Sort the result by "body\_mass\_g" in descending order.
- IV. Select the first 10 rows of penguins where "body\_mass\_g" is greater than the median body mass of all penguins.
- V. Find rows where penguins' "bill\_length\_mm" is either below 35 or above 50, and display their "species" and "sex".
- VI. Identify duplicate entries in the dataset based on the "species" and "island" columns, and remove them permanently.
- VII. Get rows for the first 5 penguins with a "body\_mass\_g" greater than 5000 and sort the result by "bill\_length\_mm".
- VIII. Replace all missing values in the "bill\_depth\_mm" column with the value that occurs most frequently in it.

- IX. Filter penguins whose "flipper\_length\_mm" is greater than 200 and "bill\_depth\_mm" is less than 18. Show only their "species", "bill\_length\_mm", and "flipper\_length\_mm".
- X. Assign ranks to penguins based on their "body\_mass\_g", with higher body masses getting higher ranks. Allow ties in the ranking.
- XI. Retrieve penguins whose "flipper\_length\_mm" is greater than 200 and "sex" is not null.
- XII. Find the rows where the "sex" column is null, and extract the "species" and "island" columns for those penguins.
- XIII. Create a new column named "Above\_Avg\_Mass" that stores True if a penguin's "body\_mass\_g" is greater than the dataset's average body mass, otherwise False.
- XIV. Extract penguins whose "island" is either "Biscoe" or "Dream" and store their information in a new DataFrame.