Question 4 / 17 - Data Science - Python 08:01 / 15:00

Instructions

We wanted to propose you an epic exercise with the famous "Iris dataset", but our csv file is garbled. What a shame!

The dataset now contains only two columns: the sepal lengths and the species. On top of that, many records were removed, so that for one value of sepal length, there is at most only one row.

Never mind, inferences can still be made, they will just be a little less accurate.

You have to implement the function `infer_iris`, which gets two parameters:

- `ref_sepal_length`: the DataFrame containing our lightweight iris dataset,
- `to_infer`: a DataFrame containing only one column "sepal length (cm)".

The function must return a DataFrame which has the same data as `to_infer`, with an additional column called "target".

For each row in `to_infer`, find the row in `ref_sepal_length` which has the nearest value of sepal length, get the corresponding target and put it in the additional column "target".

Answer

Test code

Console output

Test my code ▶ Next question