Iris

Introduction:

This exercise may seem a little bit strange, but keep doing it.

Step 1. Import the necessary libraries

```
import pandas as pd
import numpy as np
```

Step 2. Import the dataset from this address.

Step 3. Assign it to a variable called iris

```
iris = pd.read_csv("iris.csv")
iris.head()
```

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa

Step 4. Create columns for the dataset

```
# 1. sepal_length (in cm)
# 2. sepal_width (in cm)
# 3. petal_length (in cm)
# 4. petal_width (in cm)
# 5. class
```

```
iris.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']
print(iris.columns)
```

```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class'], dtype='object')
```

Step 5. Is there any missing value in the dataframe?

```
print(iris.isnull().sum())
```

```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
class           0
dtype: int64
```

Step 6. Lets set the values of the rows 10 to 29 of the column 'petal_length' to NaN

```
iris.loc[10:29, 'petal_length'] = np.nan
```

Step 7. Good, now lets substitute the NaN values to 1.0

```
iris['petal_length'].fillna(1.0, inplace=True)
```

```
<ipython-input-7-5d84e6329267>:1: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through
The behavior will change in pandas 3.0. This inplace method will never work because the intermediate object on which we

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[


  iris['petal_length'].fillna(1.0, inplace=True)
```

Step 8. Now let's delete the column class

```
iris.drop(columns='class', inplace=True)
```

Step 9. Set the first 3 rows as NaN

```
iris.iloc[0:3] = np.nan
```

Step 10. Delete the rows that have NaN

```
iris.dropna(inplace=True)
```

Step 11. Reset the index so it begins with 0 again

```
iris.reset_index(drop=True, inplace=True)
print(iris.head())
```

```
   sepal_length  sepal_width  petal_length  petal_width
0           5.0          3.6           1.4          0.2
1           5.4          3.9           1.7          0.4
2           4.6          3.4           1.4          0.3
3           5.0          3.4           1.5          0.2
4           4.4          2.9           1.4          0.2
```

BONUS: Create your own question and answer it.

Làm thế nào tôi có thể tính toán chiều dài lá đài trung bình cho từng loài trong tập dữ liệu Iris gốc?

```
iris_original = pd.read_csv("iris.csv")
iris_original.columns = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'class']


average_sepal_length = iris_original.groupby('class')['sepal_length'].mean()
print(average_sepal_length)
```

```
class
Iris-setosa        5.004082
Iris-versicolor    5.936000
Iris-virginica     6.588000
Name: sepal_length, dtype: float64
```