03_ml_03: Clustering

This dataset includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom drawn from The Audubon Society Field Guide to North American Mushrooms (1981).

Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the credibility of a mushroom; no rule like "leaflets three, let it be" for Poisonous Oak and Ivy.

Problems

Please complete class **Clustering** using the provided code template. The details are as follows:

1. Load 'ModifiedEdibleMushroom.csv' data from the "Attachment" (note: this data set has been preliminarily prepared.).
2. Choose edible mushrooms only.
3. Only the variables below have been selected to describe the distinctive characteristics of edible mushrooms: 'cap-color-rate', 'stalk-color-above-ring-rate'
4. Provide a proper data preprocessing as follows:
   - Fill missing with mean.
   - Standardize variables with Standard Scaler.
5. K-means clustering with 5 clusters (n_clusters=5, random_state=0, n_init='auto').
6. Show the maximum centroid of 2 features ('cap-color-rate' and 'stalk-color-above-ring-rate') in 2 digits.
7. Convert the centroid value to the original scale, and show the minimum centroid of 2 features in 2 digits..

And return the output based on the question number:

- For Q1, following step 1-4, please returns a data shape after Standardize variables.
- For Q2, following step 5-6, please the maximum centroid value of 2 features in 2 digits.
- For Q3, following step 7, please return the minimum original centroid scale of 2 features in 2 digits.

Submission: ** When submitting to the grader, submit **ONLY libraries, class Clustering with your modified functions**.**

Expected Results

Input	Output
Q1	(2104, 2)
Q2	[2.51 2.3]
Q3	[1.01 1.]

Template codes

```
class Clustering:
    def __init__(self, file_path): # DO NOT modify this line
        #add other parameters if needed
        self.file_path = file_path
        self.df = None #use this parameter for loading csv

    def Q1(self): # DO NOT modify this line
        """
        Step1-4
            1. Load the CSV file
            2. Choose edible mushrooms only
            3. Only the variables below have been selected to describe the distinctive
        characteristics of edible mushrooms:
                'cap-color-rate','stalk-color-above-ring-rate'
            4. Provide a proper data preprocessing as follows:
                - Fill missing with mean
                - Standardize variables with Standard Scaler
        """
        # remove pass and replace with you code
        pass

    def Q2(self): # DO NOT modify this line
        """
        Step5-6
            5. K-means clustering with 5 clusters
               (n_clusters=5, random_state=0, n_init='auto')
            6. Show the maximum centroid of 2 features ('cap-color-rate' and 'stalk-
               color-above-ring-rate') in 2 digits.
        """
        # remove pass and replace with you code
        pass

    def Q3(self): # DO NOT modify this line
        """
        Step7
            7. Convert the centroid value to the original scale, and show the minimum
               centroid of 2 features in 2 digits.
        """
        # remove pass and replace with you code
        pass
```