

Partie I : Questions de compréhension

Q1. Énumérez les quatre principales techniques d'apprentissage machine

1. Apprentissage supervisé
2. Apprentissage non supervisé
3. Apprentissage semi-supervisé
4. Apprentissage par renforcement

Q2. Face à un problème d'apprentissage machine, comment pouvons-nous anticiper s'il s'agira d'utiliser une technique d'apprentissage supervisée ou une technique d'apprentissage non-supervisée ?

Pour anticiper si une technique d'apprentissage supervisée ou non-supervisée sera utilisée, il faut analyser les données disponibles et l'objectif de l'analyse.

- Si les données sont étiquetées, une technique d'apprentissage supervisée est généralement appropriée.
- Si les données ne sont pas étiquetées et que l'objectif est de découvrir des structures ou des relations dans les données, l'apprentissage non-supervisée sera plus adaptée.

Q3. Quelles sont les étapes typiques dans un projet d'apprentissage machine ?

1. Analyses
2. Collecte et préparation des données.
3. Nettoyage des données.
4. Exploration et modélisation des données.
5. Évaluation du modèle.
6. Déploiement et maintenance du modèle.

Q4. Qu'est qu'une modalité ?

- Une modalité est une valeur ou une catégorie possible d'une variable. Par exemple, pour une variable "Genre", les modalités peuvent être "Masculin" et "Féminin".

Q5. Le coefficient d'asymétrie d'une distribution est nul. Que peut-on dire de la distribution ?

- Si le coefficient d'asymétrie d'une distribution est nul, cela signifie que la distribution est symétrique par rapport à sa moyenne. Les valeurs sont réparties de manière égale de chaque côté de la moyenne.

Partie II : Description des données

1. Charger les données

```
In [5]: import pandas as pd

df = pd.read_excel("C:\\Users\\bouma12\\VSCode_WorkSpace\\Fonction_travail\\examen_
print("Premières lignes du DataFrame:")
print(df.head())
```

Premières lignes du DataFrame:

	Individu	Odor	Fruity	Flower	Aroma /intensity	Aroma /persistency	\
0	2EL	3.407	2.885	2.320	3.259		2.963
1	1CHA	3.370	2.560	2.440	2.962		2.808
2	1FON	3.250	2.769	2.192	3.077		2.800
3	1VAU	3.160	2.391	2.083	2.542		2.583
4	1DAM	3.536	3.160	2.231	3.615		3.296

	Aroma /quality	Spice	Acidity	Balance	Smooth	Intensity	Harmony	\
0	3.200	1.840	2.107	3.250	2.731	2.857	3.143	
1	2.926	1.739	2.107	2.926	2.500	2.893	2.964	
2	3.077	2.250	2.179	3.321	2.679	3.074	3.143	
3	2.478	2.167	3.179	2.333	1.680	2.462	2.038	
4	3.462	2.148	2.571	3.464	3.036	3.643	3.643	

	Region	Soil
0	Saumur	Env1
1	Saumur	Env1
2	Bourgueuil	Env1
3	Chinon	Env2
4	Saumur	Reference

```
C:\Users\bouma12\AppData\Roaming\Python\Python312\site-packages\openpyxl\worksheet\
reader.py:329: UserWarning: Unknown extension is not supported and will be removed
warn(msg)
```

2. Déterminez le nombre d'individus et de variables et identifiez le type des variables

```
In [ ]: num_individuals = df.shape[0]
num_variables = df.shape[1]

# Types of variables
variable_types = df.dtypes

print(f"Nombre des individus: {num_individuals}")
print(f"Nombre des variables: {num_variables}")

print("Types des variables:")
print(variable_types)
```

Nombre des individus: 21

Nombre des variables: 15

Premières lignes du DataFrame:

	Individu	Odor	Fruity	Flower	Aroma /intensity	Aroma /persistency	\
0	2EL	3.407	2.885	2.320	3.259		2.963
1	1CHA	3.370	2.560	2.440	2.962		2.808
2	1FON	3.250	2.769	2.192	3.077		2.800
3	1VAU	3.160	2.391	2.083	2.542		2.583
4	1DAM	3.536	3.160	2.231	3.615		3.296

	Aroma /quality	Spice	Acidity	Balance	Smooth	Intensity	Harmony	\
0	3.200	1.840	2.107	3.250	2.731	2.857	3.143	
1	2.926	1.739	2.107	2.926	2.500	2.893	2.964	
2	3.077	2.250	2.179	3.321	2.679	3.074	3.143	
3	2.478	2.167	3.179	2.333	1.680	2.462	2.038	
4	3.462	2.148	2.571	3.464	3.036	3.643	3.643	

	Region	Soil
0	Saumur	Env1
1	Saumur	Env1
2	Bourgueuil	Env1
3	Chinon	Env2
4	Saumur	Reference

Types des variables:

Individu	object
Odor	float64
Fruity	float64
Flower	float64
Aroma /intensity	float64
Aroma /persistency	float64
Aroma /quality	float64
Spice	float64
Acidity	float64
Balance	float64
Smooth	float64
Intensity	float64
Harmony	float64
Region	object
Soil	object

dtype: object

3. Calculez la moyenne et l'écart-type de la variable Intensity

```
In [7]: moyenne_intensity = df['Intensity'].mean()
sigma_intensity = df['Intensity'].std()

print(f"Moyenne de la variable Intensity: {moyenne_intensity:.2f}")
print(f"Écart-type de la variable Intensity: {sigma_intensity:.2f}")
```

Moyenne de la variable Intensity: 3.17

Écart-type de la variable Intensity: 0.37

4. Déterminez les modalités de la variable Region

```
In [8]: modalites_region = df['Region'].unique()
print("Modalités de la variable Region:")
print(modalites_region)
```

```
Modalités de la variable Region:
['Saumur' 'Bourgueuil' 'Chinon']
```

5. Calculez l'effectif de chaque modalité de Region

```
In [9]: effectif_region = df['Region'].value_counts()
print("Effectif de chaque modalité de Region:")
print(effectif_region)
```

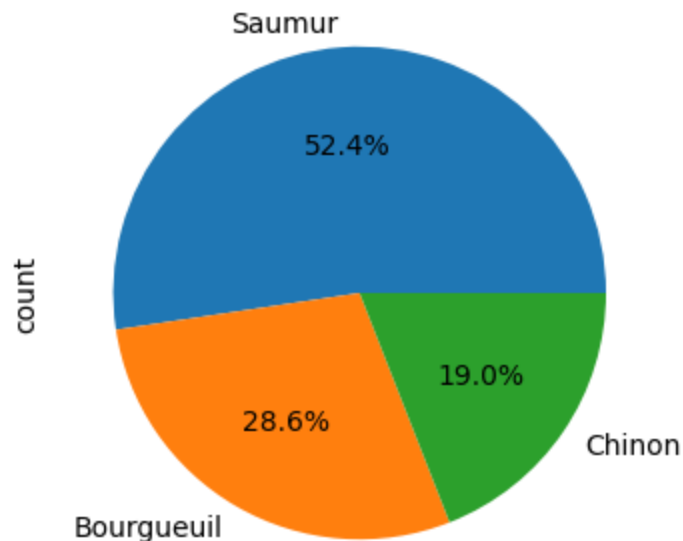
```
Effectif de chaque modalité de Region:
Region
Saumur      11
Bourgueuil   6
Chinon       4
Name: count, dtype: int64
```

6. Visualiser le diagramme circulaire de la variable Region

```
In [19]: effectif_region.plot.pie(
    autopct='%1.1f%%',
    figsize=(4, 4),
    title="Répartition des modalités de la variable Region"
)
```

```
Out[19]: <Axes: title={'center': 'Répartition des modalités de la variable Region'}, ylabel='count'>
```

Répartition des modalités de la variable Region



7. Déterminez les caractéristique de tendance centrale de la variable Spice

```
In [20]: moyenne_spice = df['Spice'].mean()
print(f"La moyenne de Spice est {moyenne_spice:.2f} ans")

mediane_spice = df['Spice'].median()
print(f"La médiane de Spice est {mediane_spice:.2f} ans")

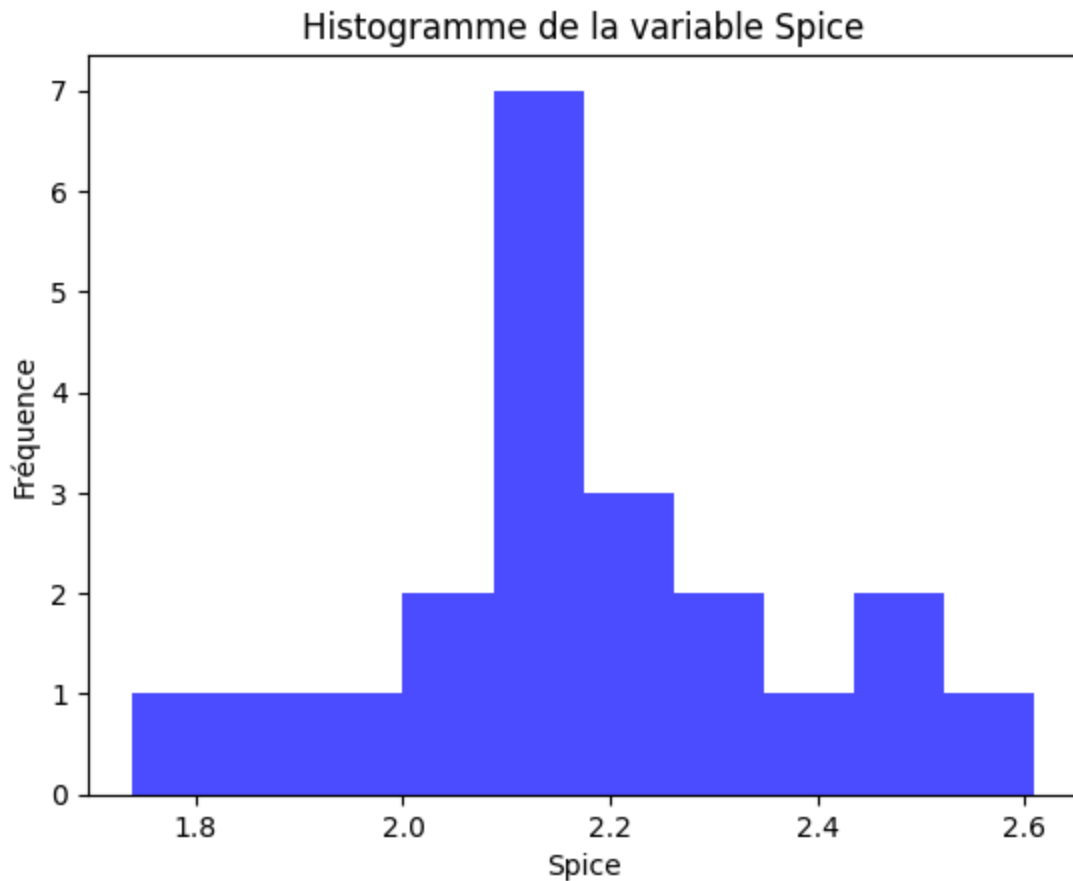
quartile_1_spice = df['Spice'].quantile(0.25)
print(f"Le premier quartile d'âge est {quartile_1_spice:.2f} ans")
quartile_2_spice = df["Spice"].quantile(0.50)
print(f"Le deuxième quartile d'âge est {quartile_2_spice:.2f} ans")
quartile_3_spice = df["Spice"].quantile(0.75)
print(f"Le troisième quartile d'âge est {quartile_3_spice:.2f} ans")
```

La moyenne de Spice est 2.18 ans
La médiane de Spice est 2.16 ans
Le premier quartile d'âge est 2.11 ans
Le deuxième quartile d'âge est 2.16 ans
Le troisième quartile d'âge est 2.28 ans

8. Représentez l'histogramme de la variable Spice

```
In [21]: import matplotlib.pyplot as plt

# Histogram of the variable Spice
df['Spice'].plot.hist(
    bins=10,
    alpha=0.7,
    color='blue',
    title="Histogramme de la variable Spice"
)
plt.xlabel("Spice")
plt.ylabel("Fréquence")
plt.show()
```



9. Calculez le coefficient d'asymétrie de la variable Spice

```
In [22]: coefficient_asymetrie = df['Spice'].skew(skipna=True)
print(f"Le coefficient d'asymétrie de la variable 'Spice' est {coefficient_asymetrie}")

# Interprétation
if coefficient_asymetrie > 0:
    print("La distribution est asymétrique à droite.")
elif coefficient_asymetrie < 0:
    print("La distribution est asymétrique à gauche.")
else:
    print("La distribution est symétrique.")
```

Le coefficient d'asymétrie de la variable 'Spice' est -0.12
La distribution est asymétrique à gauche.

10. Interprétez les résultats

La distribution est asymétrique à gauche.