

# Week 3 Quiz

Zichen - zz3119

## Instructions

Replace the Name and UNI in cell above and the notebook filename

Replace all '\_\_\_\_' below using the instructions provided.

When completed,

1. make sure you've replaced Name and UNI in the first cell and filename
2. Kernel -> Restart & Run All to run all cells in order
3. Print Preview -> Print (Landscape Layout) -> Save to pdf
4. post pdf to GradeScope

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set_style('darkgrid')

%matplotlib inline
```

```
In [2]: # Use pandas to read in 'wine_dataset.csv'
# This is a dataset of various wines with a target of categorical variable '
df = pd.read_csv('../data/wine_dataset.csv')
```

```
In [3]: # 1. Print out the number of rows and columns in the dataset using .shape
df.shape
```

Out[3]: (178, 14)

```
In [4]: # 2. Display the first 3 rows of df using .head()
df.head(3)
```

```
Out[4]:
```

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24

```
In [5]: # 3. Display the summary stats of numeric columns using .describe()
df.describe()
```

Out [5]:

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols
<b>count</b>	178.000000	178.000000	178.000000	178.000000	178.000000	178.0000
<b>mean</b>	13.000618	2.336348	2.366517	19.494944	99.741573	2.2951
<b>std</b>	0.811827	1.117146	0.274344	3.339564	14.282484	0.6258
<b>min</b>	11.030000	0.740000	1.360000	10.600000	70.000000	0.9800
<b>25%</b>	12.362500	1.602500	2.210000	17.200000	88.000000	1.7425
<b>50%</b>	13.050000	1.865000	2.360000	19.500000	98.000000	2.3550
<b>75%</b>	13.677500	3.082500	2.557500	21.500000	107.000000	2.8000
<b>max</b>	14.830000	5.800000	3.230000	30.000000	162.000000	3.8800

In [6]: *# 4. Using .iloc[], display the first 3 rows, first 3 columns*  
*# You should see the columns ['alcohol','malic\_acid','ash']*  
`df.iloc[:3, :3]`

Out [6]:

	alcohol	malic_acid	ash
<b>0</b>	14.23	1.71	2.43
<b>1</b>	13.20	1.78	2.14
<b>2</b>	13.16	2.36	2.67

In [7]: *# 5. Using .loc[], display rows with index label 4 to 6 inclusive and columns*  
`df.loc[4:6, ['ash', 'total_phenols']]`

Out [7]:

	ash	total_phenols
<b>4</b>	2.87	2.80
<b>5</b>	2.45	3.27
<b>6</b>	2.45	2.50

In [8]: *# 6. Return the 'ash' and 'hue' columns for all rows with 'hue' greater than*  
*# This should result in 89 rows x 2 columns*  
*# Note that pandas will only display a subset of the rows*  
`df[df['hue'] > df['hue'].median()][['ash', 'hue']]`

Out [8]:

	ash	hue
0	2.43	1.04
1	2.14	1.05
2	2.67	1.03
4	2.87	1.04
5	2.45	1.05
...	...	...
112	2.92	1.23
113	2.50	1.10
115	2.20	1.71
117	2.19	1.06
127	2.78	0.97

89 rows × 2 columns

```
In [9]: # 7.1. Create two axes using plt.subplots with 1 row , 2 columns, figsize=(10, 4)
fig, ax = plt.subplots(1, 2, figsize=(10, 4))

# 7.2 In the first axis (ax[0]), plot the distribution of df.alcohol using sns.histplot
sns.histplot(data=df, x='alcohol', ax=ax[0])

# 7.3 Add the title 'Wine Alcohol Content' to ax[0] using .set_title()
ax[0].set_title('Wine Alcohol Content')

# 7.4 In the second axis, plot a boxplot of df.color_intensity using sns.boxplot
sns.boxplot(data=df, x='color_intensity', ax=ax[1])

# 7.5 Add the title 'Wine Color Intensity' to ax[1] using .set_title()
ax[1].set_title('Wine Color Intensity')
plt.tight_layout()
plt.show()
```

