Week 3 Quiz

Clarissa Tai - rt2822

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 'class'
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	nonflavanoid_phenols	proanthocyanins	color_intensity
0	14.23	1.71	2.43	15.6	127.0	2.80	3.06	0.28	2.29	5.64
1	13.20	1.78	2.14	11.2	100.0	2.65	2.76	0.26	1.28	4.38
2	13.16	2.36	2.67	18.6	101.0	2.80	3.24	0.30	2.81	5.68

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	flavanoids	nonflavanoid_phenols	proanthocyanins
count	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000	178.000000
mean	13.000618	2.336348	2.366517	19.494944	99.741573	2.295112	2.029270	0.361854	1.590899
std	0.811827	1.117146	0.274344	3.339564	14.282484	0.625851	0.998859	0.124453	0.572359
min	11.030000	0.740000	1.360000	10.600000	70.000000	0.980000	0.340000	0.130000	0.410000
25%	12.362500	1.602500	2.210000	17.200000	88.000000	1.742500	1.205000	0.270000	1.250000
50%	13.050000	1.865000	2.360000	19.500000	98.000000	2.355000	2.135000	0.340000	1.555000
75%	13.677500	3.082500	2.557500	21.500000	107.000000	2.800000	2.875000	0.437500	1.950000
max	14.830000	5.800000	3.230000	30.000000	162.000000	3.880000	5.080000	0.660000	3.580000

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
0	14.23	1.71	2.43
1	13.20	1.78	2.14
2	13.16	2.36	2.67

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

Out[7]:

	ash	total_phenols
4	2.87	2.80
5	2.45	3.27
6	2.45	2.50

In [8]:
```
# 6. Return the 'ash' and 'hue' columns for all rows with 'hue' greater than the median value of hue
# 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','total_phenols']]
```

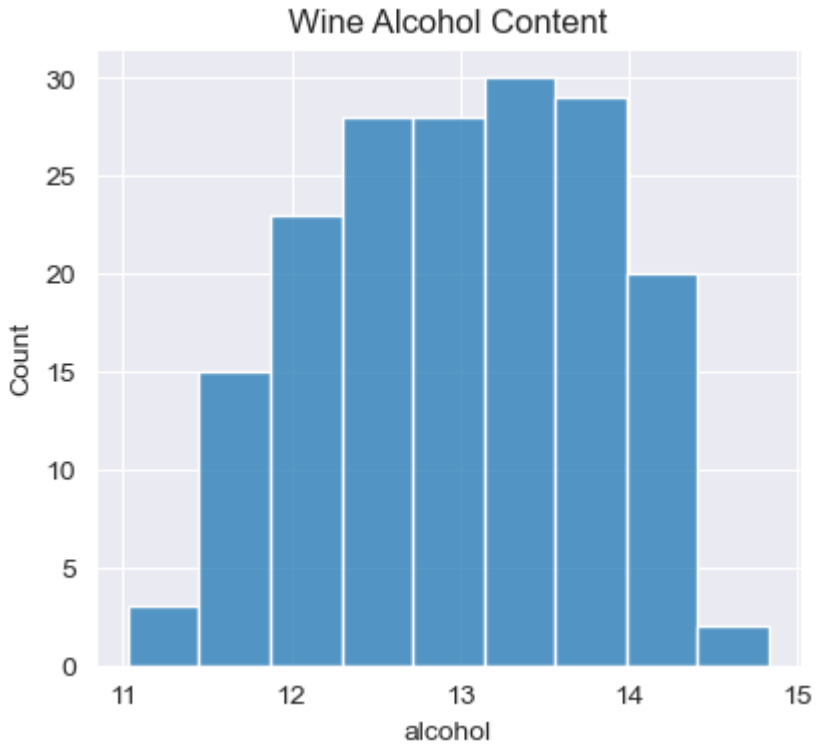
Out[8]:

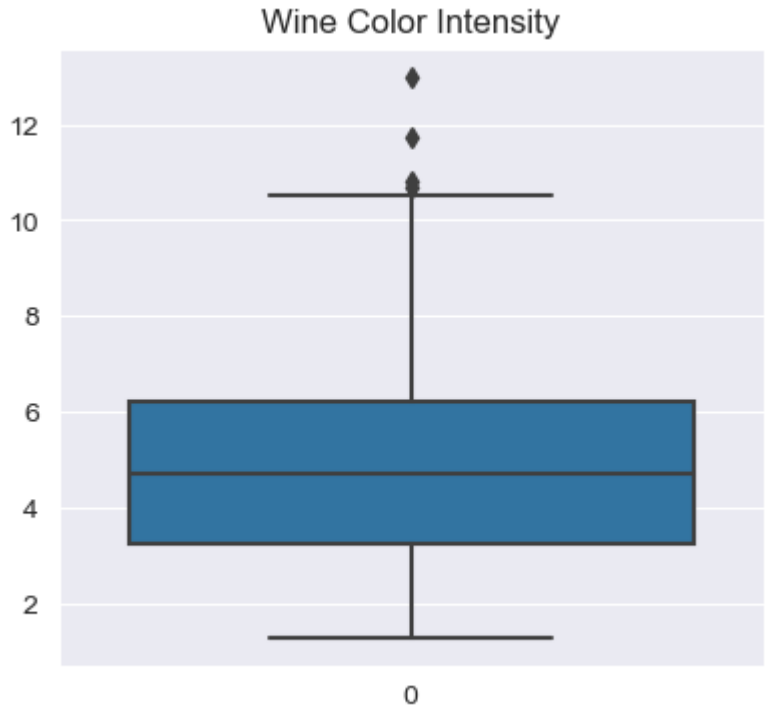
	ash	total_phenols
0	2.43	2.80
1	2.14	2.65
2	2.67	2.80
4	2.87	2.80
5	2.45	3.27
...	...	...
112	2.92	1.75
113	2.50	2.48
115	2.20	2.46
117	2.19	2.00
127	2.78	2.13

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(ax=ax[0], data = df.alcohol)

# 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(ax=ax[1], data = df.color_intensity)

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

Out[9]:
```
Text(0.5, 1.0, 'Wine Color Intensity')
```


Wine Alcohol Content
Count
alcohol
0
5
10
15
20
25
30
11
12
13
14
15


Wine Color Intensity
2
4
6
8
10
12
0

In [ ]: