1

January 6, 2021

[1]:
```
from utils import get_dataset, col_type, COL_TYPE, print_dict
import matplotlib.pyplot as plt
import pprint
```

[2]:
```
dataset = get_dataset('dataset.xlsx')
```

0.0.1 Question 1

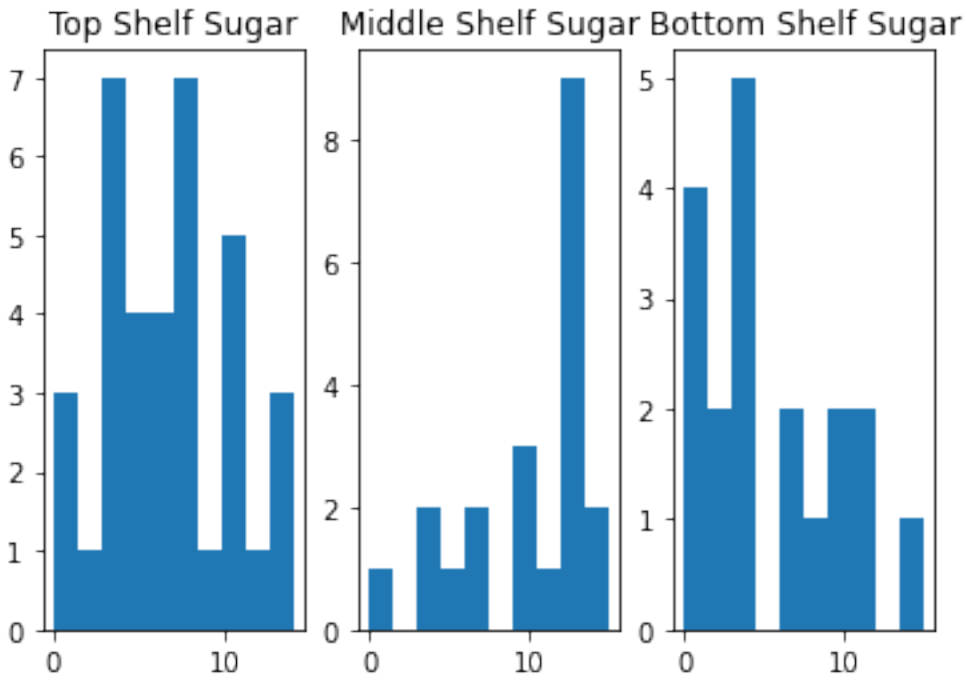
[3]:
```
quant_count = 0
qual_count = 0
for i in dataset:
    if(col_type(dataset[i]) == COL_TYPE.QUALITATIVE):
        qual_count += 1
    elif(col_type(dataset[i]) == COL_TYPE.QUANTITATIVE):
        quant_count += 1
print("==> Quantitative Column Count: ", quant_count)
print("==> Qualitative Column Count: ", qual_count)
```

```
==> Quantitative Column Count:  11
==> Qualitative Column Count:  4
```

0.0.2 Question 2

[4]:
```
top_shelf = dataset[(dataset['shelf'] == 'Top')]
middle_shelf = dataset[(dataset['shelf'] == 'Middle')]
bottom_shelf = dataset[(dataset['shelf'] == 'Bottom')]
fig, axs = plt.subplots(1, 3)

axs[0].hist(top_shelf['sugars'])
axs[0].set_title('Top Shelf Sugar')

axs[1].hist(middle_shelf['sugars'])
axs[1].set_title('Middle Shelf Sugar')

axs[2].hist(bottom_shelf['sugars'])
axs[2].set_title('Bottom Shelf Sugar')
```

```
mean = {'top': top_shelf['sugars'].mean(), 'mid': middle_shelf['sugars'].
 ↪mean(), 'bot': bottom_shelf['sugars'].mean()}

plt.show()
```



0.0.3 Question 3

Order of shelf with higest sugar content.

[5]:
```
print_dict(dict(sorted(mean.items(), key=lambda item: item[1], reverse=True)))
```

```
mid (9.619047619047619)
top (6.527777777777778)
bot (5.105263157894737)
```

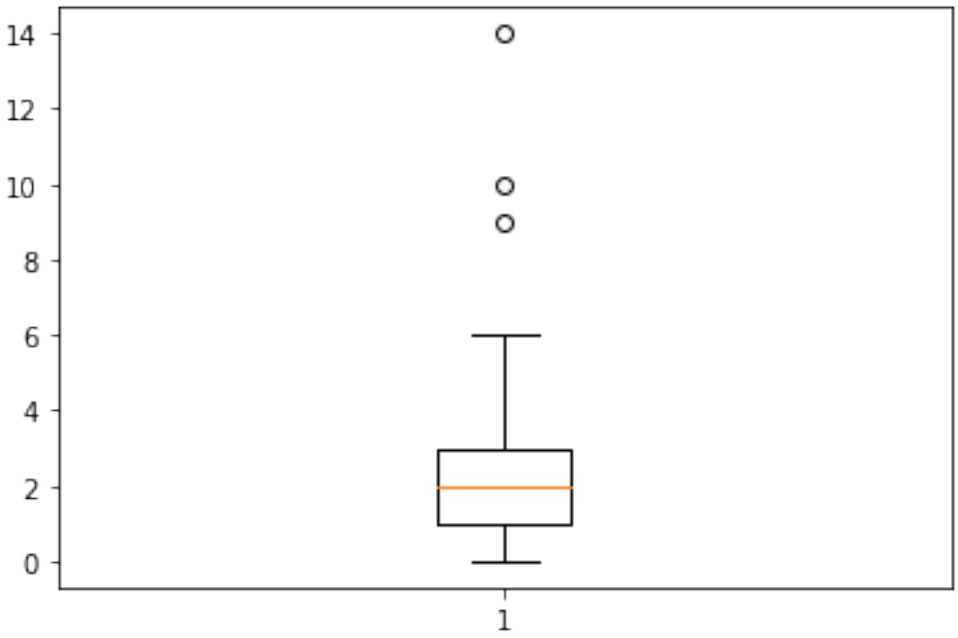
0.0.4 Question 4

Five-number summary plot for the variable 'fiber'

[6]:
```
plt.boxplot(dataset['fiber'])
dataset['fiber'].describe()
```

[6]:
```
count    77.000000
mean      2.151948
std       2.383364
min       0.000000
```

```
25%       1.000000
50%       2.000000
75%       3.000000
max      14.000000
Name: fiber, dtype: float64
```

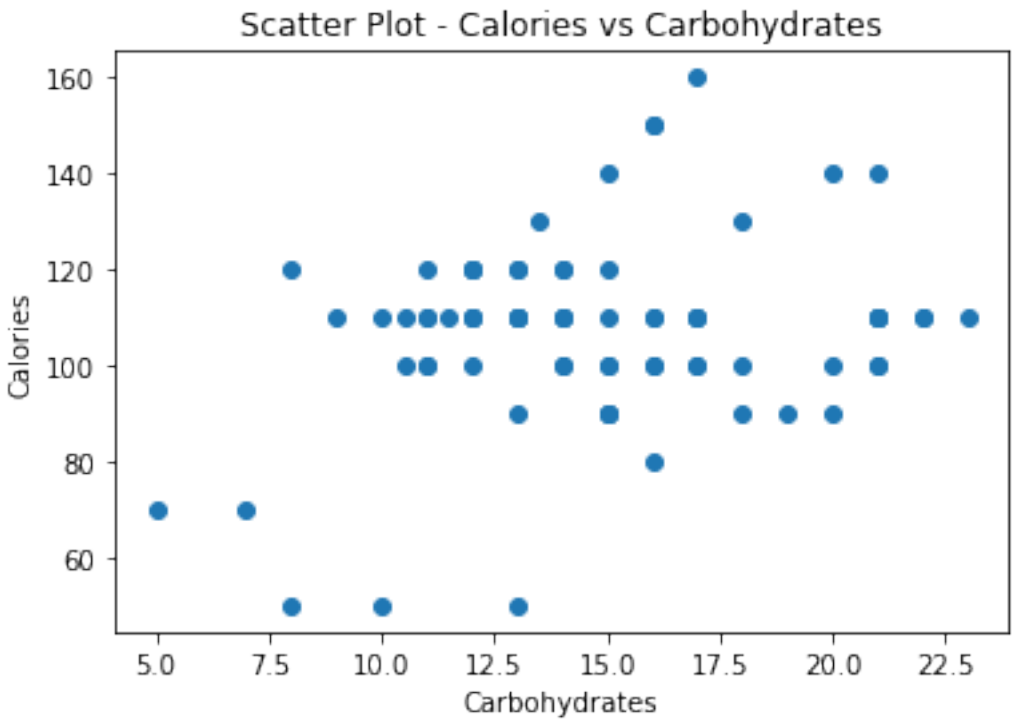


0.0.5 Question 5

Scatter plot between calories and carbohydrates

[7]:
```
plt.scatter(dataset['carbo'], dataset['calories'])
plt.title('Scatter Plot - Calories vs Carbohydrates')
plt.ylabel('Calories')
plt.xlabel('Carbohydrates')
dataset['carbo'].corr(dataset['calories'])
```

[7]:
```
0.25763783073781144
```


Scatter Plot - Calories vs Carbohydrates
Calories
160
140
120
100
80
60
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
Carbohydrates