Importing the Header files

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

Reading the file

In [2]:
```
df= pd.read_csv("Coffee.csv")
df.head(50)
```

Out[2]:

	CodingHours	CoffeeCupsPerDay	CoffeeTime	CodingWithoutCoffee
0	8	2	Before coding	Yes
1	3	2	Before coding	Yes
2	5	3	While coding	No
3	8	2	Before coding	No
4	10	3	While coding	Sometimes
5	8	2	While coding	Sometimes
6	5	2	While coding	Yes
7	10	4	Before coding	Sometimes
8	10	2	While coding	Yes
9	10	2	While coding	Yes
10	10	3	While coding	Sometimes
11	2	3	Before coding	Sometimes
12	8	2	Before and while coding	No
13	9	3	While coding	Sometimes
14	6	1	Before coding	Yes
15	6	3	While coding	No
16	10	3	While coding	Sometimes
17	6	4	While coding	Yes
18	8	3	While coding	No
19	3	1	In the morning	Yes
20	3	3	While coding	Yes
21	9	5	While coding	Sometimes
22	4	2	While coding	Sometimes

	CodingHours	CoffeeCupsPerDay	CoffeeTime	CodingWithoutCoffee
23	8	3	While coding	No
24	5	3	While coding	Sometimes
25	6	3	While coding	Yes
26	7	4	Before coding	Sometimes
27	8	4	While coding	Sometimes
28	7	1	Before coding	Yes
29	2	2	While coding	Sometimes
30	10	6	All the time	No
31	2	2	While coding	No
32	9	2	Before coding	Sometimes
33	4	1	Before coding	Yes
34	6	2	While coding	Sometimes
35	2	2	While coding	Yes
36	2	3	All the time	Sometimes
37	7	7	After coding	Yes
38	1	1	While coding	Yes
39	4	2	While coding	Yes
40	8	2	While coding	Sometimes
41	8	4	While coding	Sometimes
42	3	2	While coding	Sometimes
43	10	6	While coding	Sometimes
44	3	2	After coding	Yes
45	7	2	While coding	Sometimes
46	2	1	While coding	Sometimes
47	9	1	In the morning	Yes
48	8	2	Before coding	No
49	6	1	Before coding	Yes

In [76]:
```
df.describe()
```

Out[76]:

	CodingHours	CoffeeCupsPerDay
count	50.000000	50.000000
mean	6.300000	2.620000
std	2.801239	1.338443

	CodingHours	CoffeeCupsPerDay
min	1.000000	1.000000
25%	4.000000	2.000000
50%	7.000000	2.000000
75%	8.000000	3.000000
max	10.000000	7.000000

In [77]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 4 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   CodingHours          50 non-null     int64 
 1   CoffeeCupsPerDay     50 non-null     int64 
 2   CoffeeTime           50 non-null     object
 3   CodingWithoutCoffee  50 non-null     object
dtypes: int64(2), object(2)
memory usage: 1.7+ KB
```

Ploting a Bar Graph with Frequency on y-axis and Coffee Time on x-axis

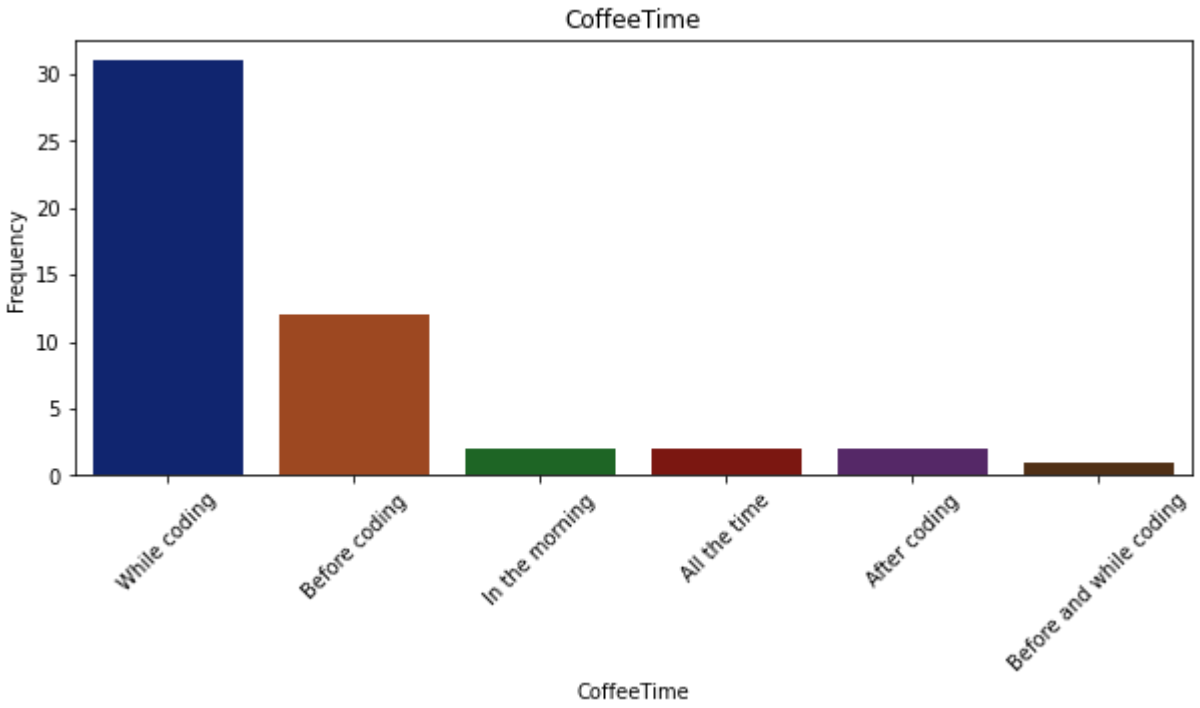
In [78]:
```
def count_plot(variable):
    """
        input: variable example: "CoffeTime"
        output: count plot and value count
    """
    # get feature
    var = df[variable]

    #visualization
    plt.figure(figsize=(10,4))
    sns.countplot(x=var, palette="dark", order=var.value_counts().index)
    plt.xticks(rotation=45)
    plt.ylabel("Frequency")
    plt.title(variable)
    print("{}".format(var.value_counts()))
    plt.show()
```

In [79]:
```
categorical = ["CoffeeTime"]
for i in categorical:
    count_plot(i)
```

```
While coding               31
Before coding              12
In the morning              2
All the time                2
After coding                2
```

```
Before and while coding     1
Name: CoffeeTime, dtype: int64
```



Inference

From the graph, we can derive that the number of times people drink coffee is high while they are coding compared to the number of times they drink coffee at other times of the day. We can assume from the graph that coffee consumption increases coding rate by providing necessary refreshment.