In [148…
```
import pandas as pd
import numpy as np
import sklearn
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler , LabelEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import LabelEncoder
import warnings
warnings.filterwarnings('ignore')
```

In [149…
```
df=pd.read_csv('tips.csv')
```

1. what is the average tip (as a percentage of meal cost) for for lunch and for dinner?

In [150…
```
df['tip_percentage']=df['tip']/df['total_bill']*100
df.groupby('time')['tip_percentage'].mean()
```

Out[150]:
```
time
Dinner    15.951779
Lunch     16.412793
Name: tip_percentage, dtype: float64
```

1. what is average tip for each day of the week (as a percentage of meal cost)?

In [151…
```
df.groupby('day')['tip_percentage'].mean()
```

Out[151]:
```
day
Fri     16.991303
Sat     15.315172
Sun     16.689729
Thur    16.127563
Name: tip_percentage, dtype: float64
```

1. when are tips highest (which day and time)?

In [152…
```
print(df.groupby(['day','time'])['tip'].max())
print("the highest tip is on sunday at dinner")
```

```
day   time
Fri   Dinner     4.73
      Lunch      3.48
Sat   Dinner    10.00
Sun   Dinner     6.50
Thur  Dinner     3.00
      Lunch      6.70
Name: tip, dtype: float64
the highest tip is on sunday at dinner
```

1. compute the correlation between meal prices and tips

In [153…
```
print("the correlation between meal prices and tips is",df['total_bill'].corr(c
```
```
the correlation between meal prices and tips is 0.6757341092113646
```

1. is there any relationship between tips and size of the group?

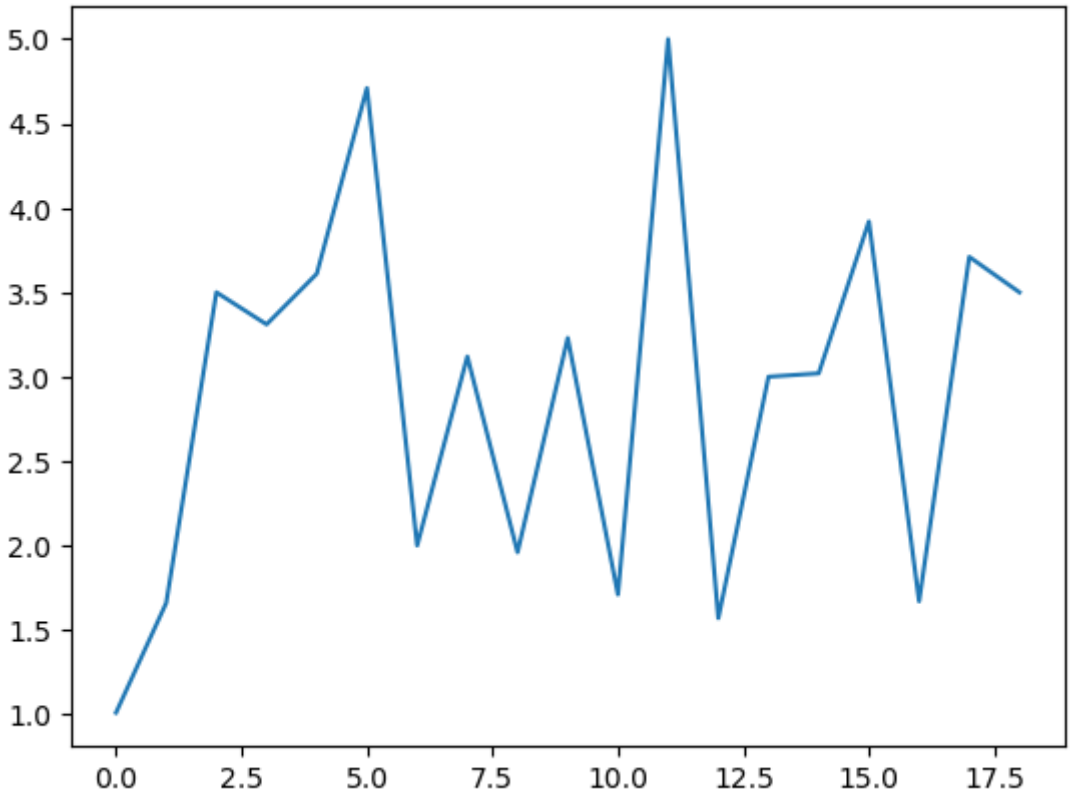
In [154…
```
# is there any relationship between tips and size of the group?
print("the correlation between meal prices and tips is",df["size"].corr(df['tip
```
```
the correlation between meal prices and tips is 0.48929877523035725
```

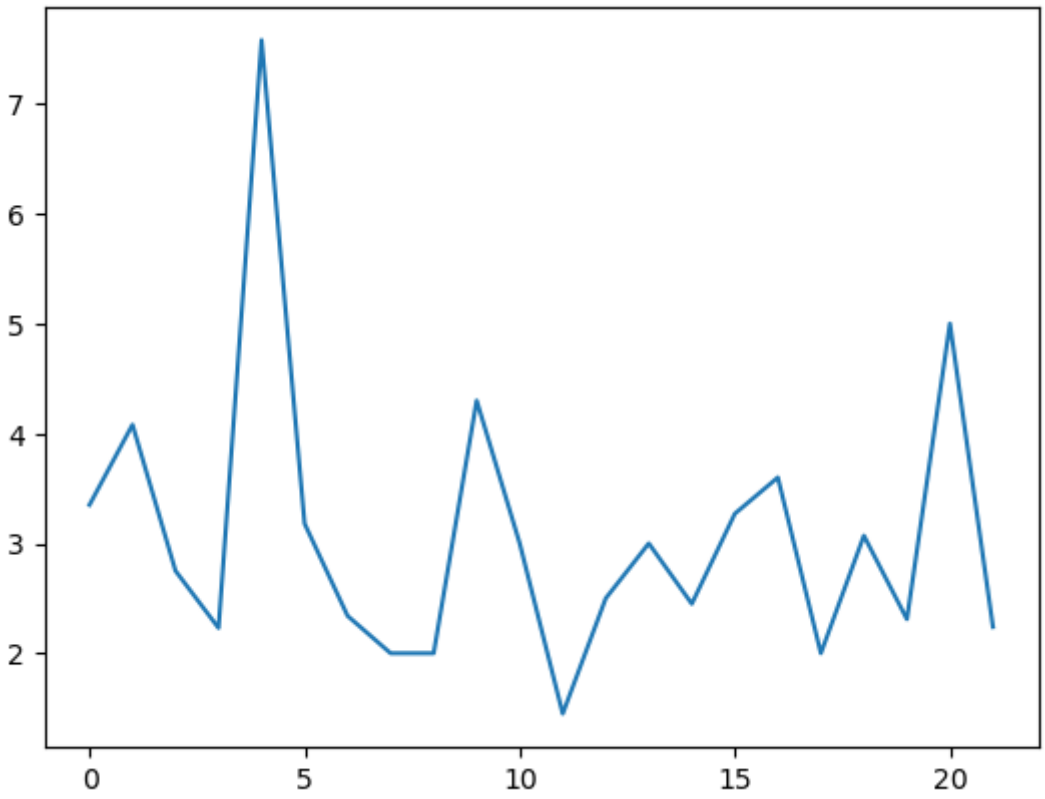
1. what percentage of people are smoking?

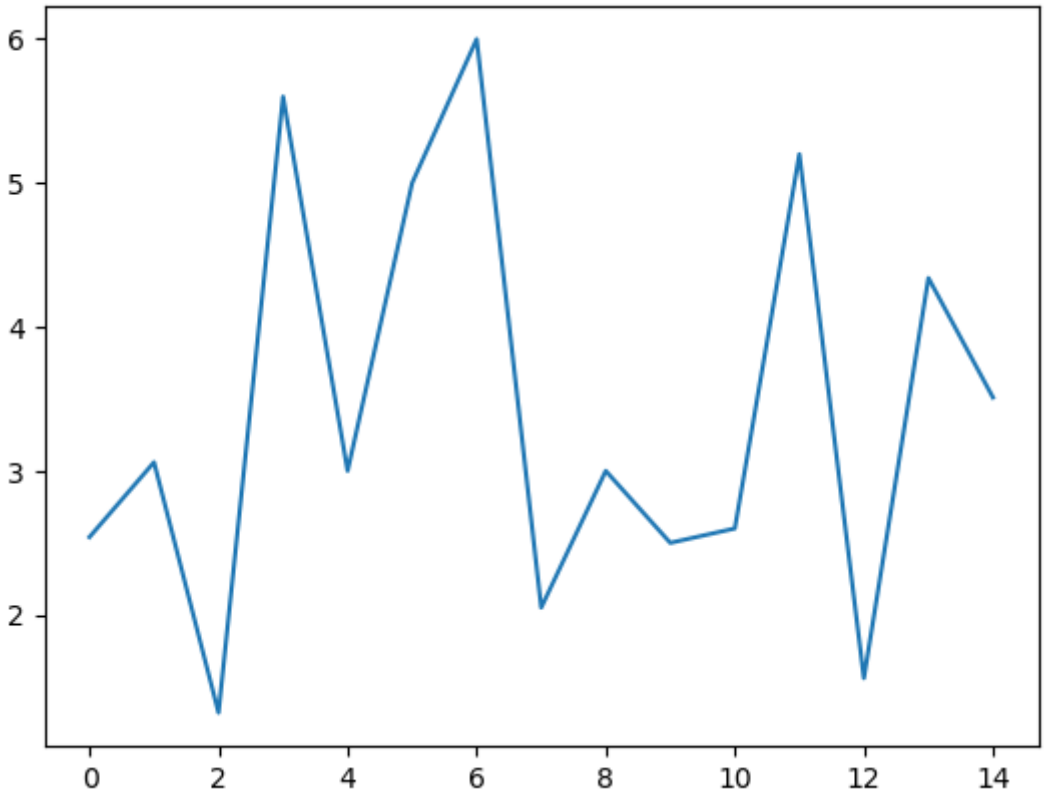
In [155…
```
#what percentage of people are smoking?
print(df['smoker'].value_counts(normalize=True)*100)
print("the percentage of people who smoke is 38.5%")
```
```
No     61.885246
Yes    38.114754
Name: smoker, dtype: float64
the percentage of people who smoke is 38.5%
```

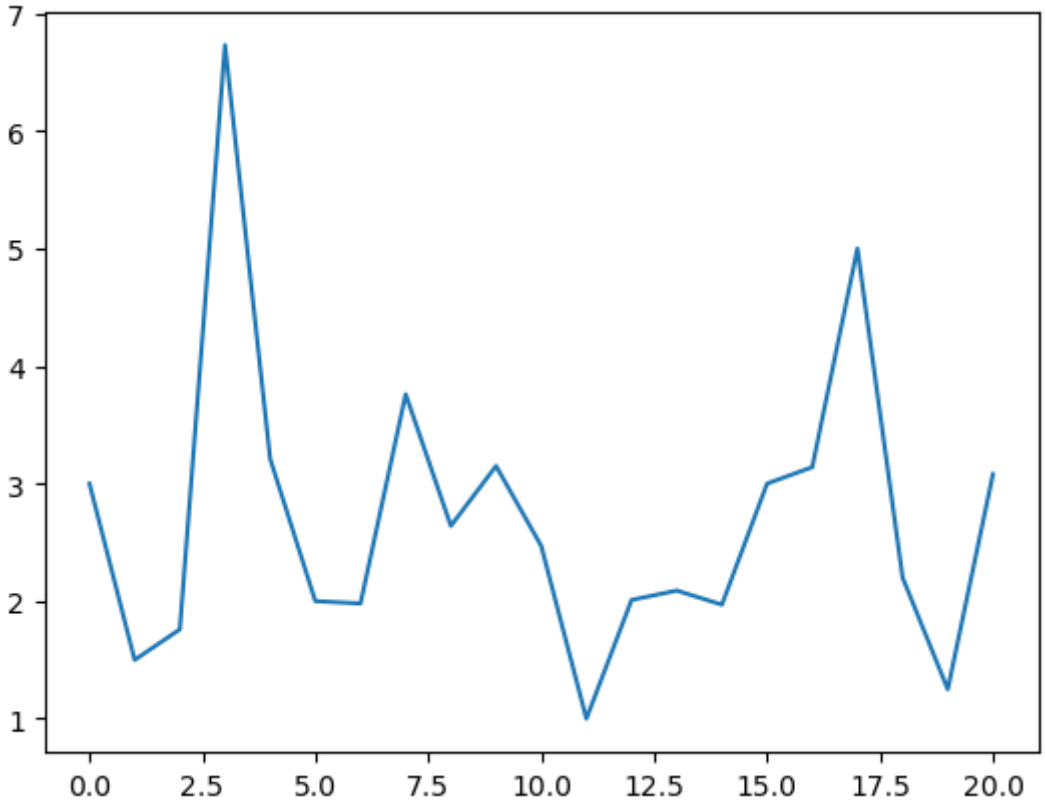
1. assume that rows in the tips.csv file are arranged in time. Are tips increasing with time in each day?

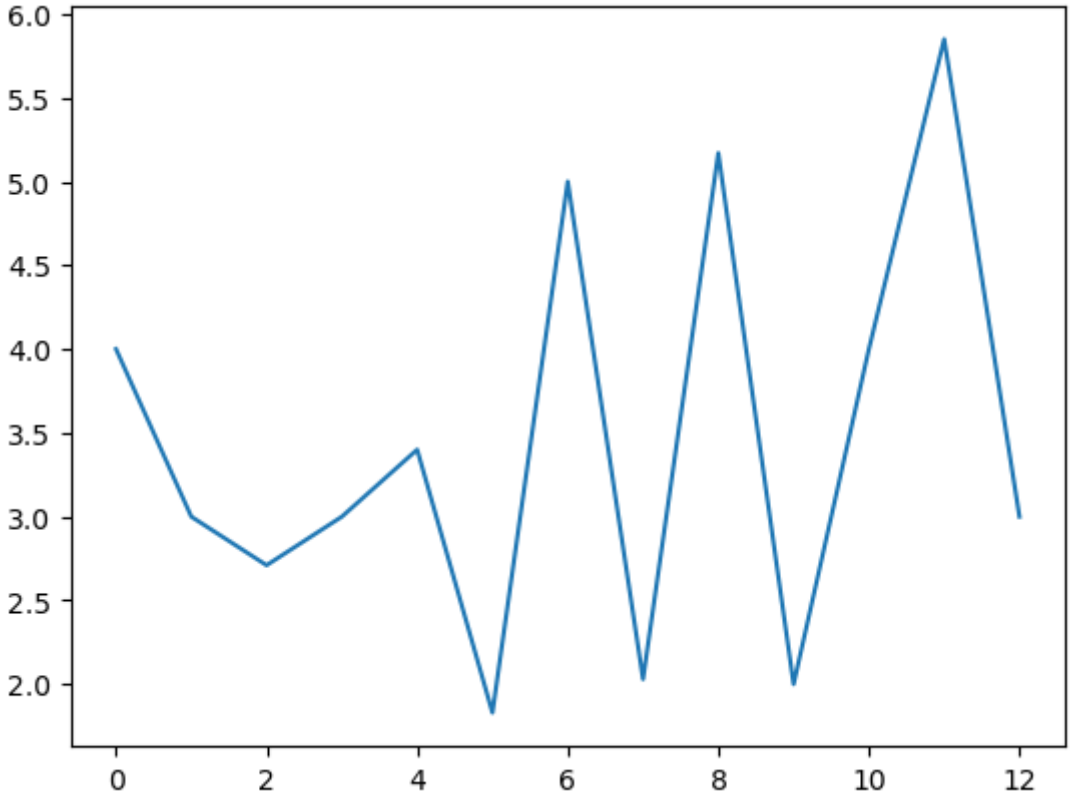
In [156…
```
# assume that rows in the tips.csv file are arranged in time. Are tips increasi
tip_arr=[]
for i in range(len(df)-1):
    tip_arr.append(df["tip"][i])
    if df["day"][i]!=df["day"][i+1]:
        plt.plot(tip_arr)
        plt.show()
        tip_arr=[]
print("we can see that tips are not increasing with time in each day, it's irre
```

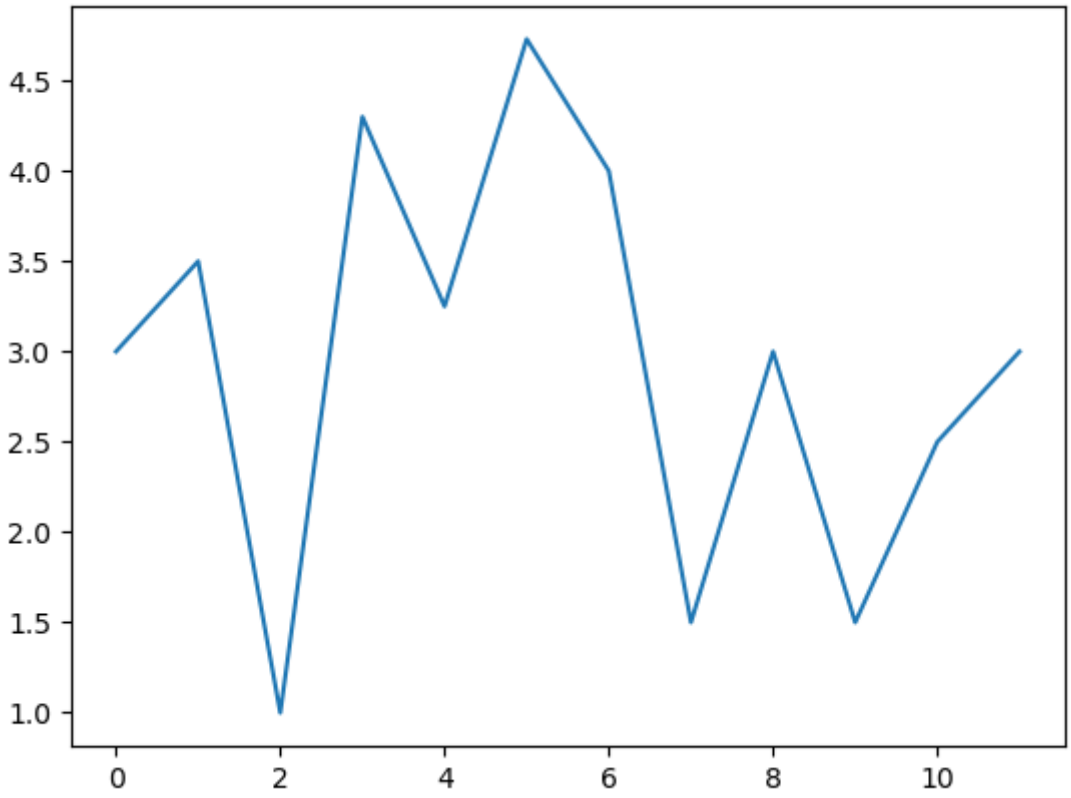

5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

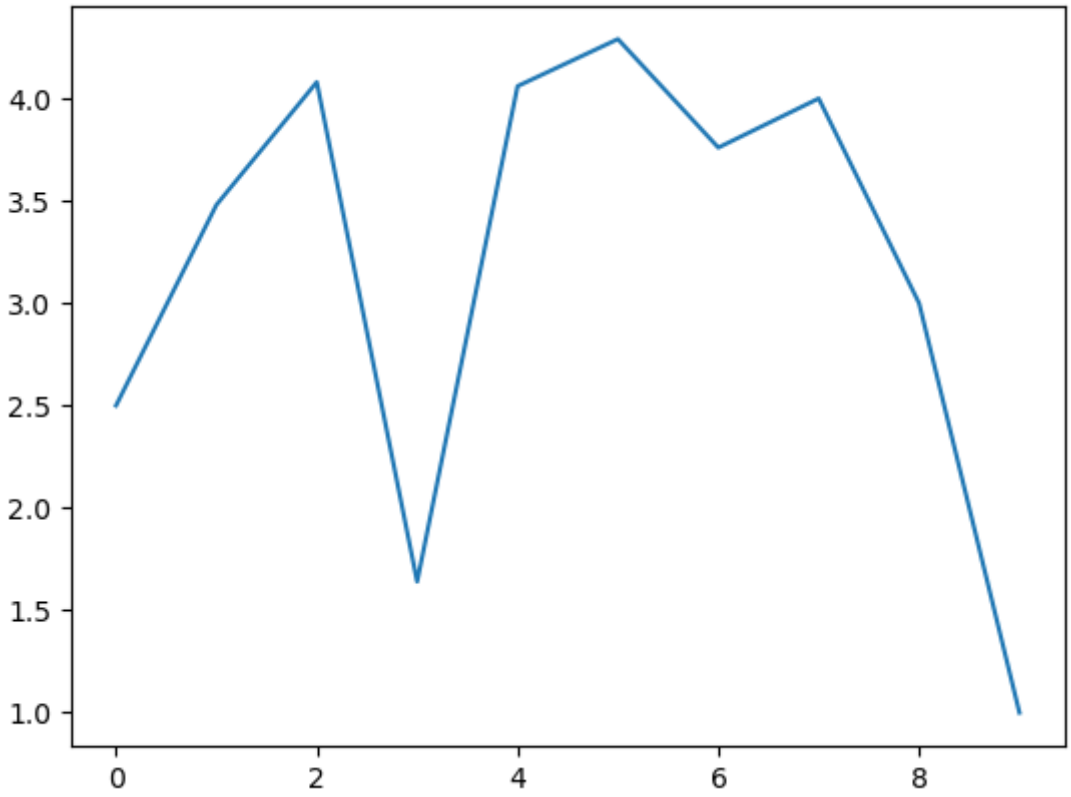

7
6
5
4
3
2
0
5
10
15
20

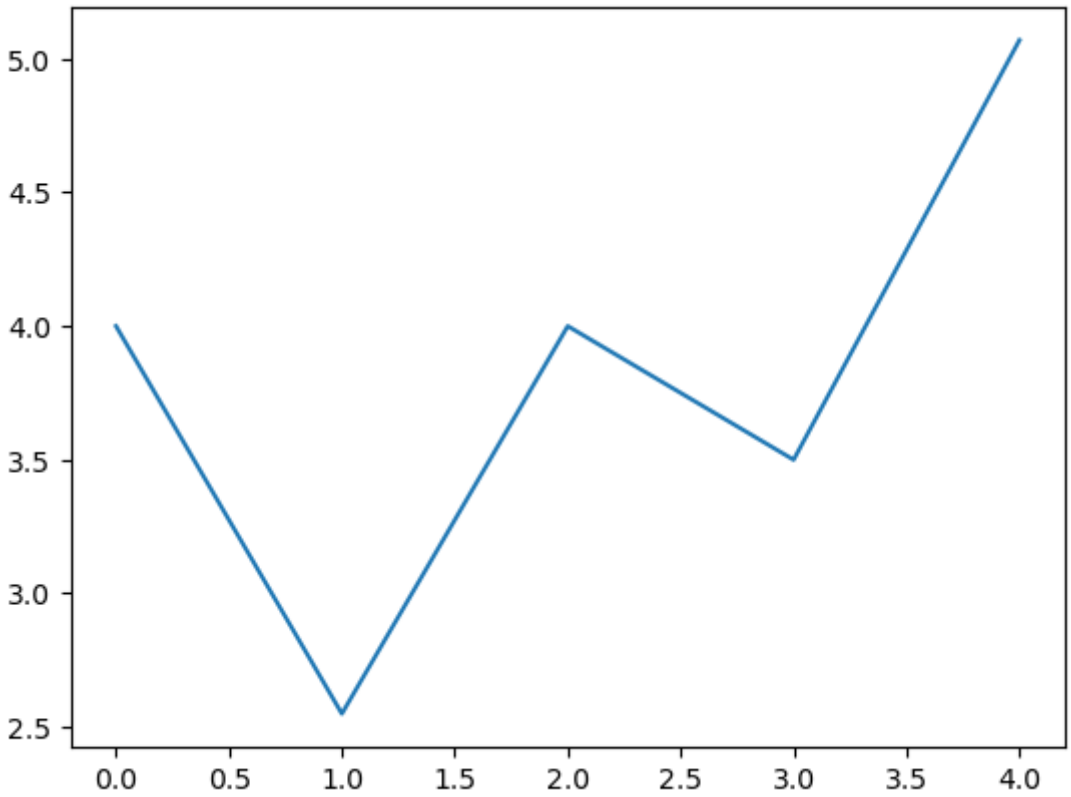

6
5
4
3
2
0
2
4
6
8
10
12
14

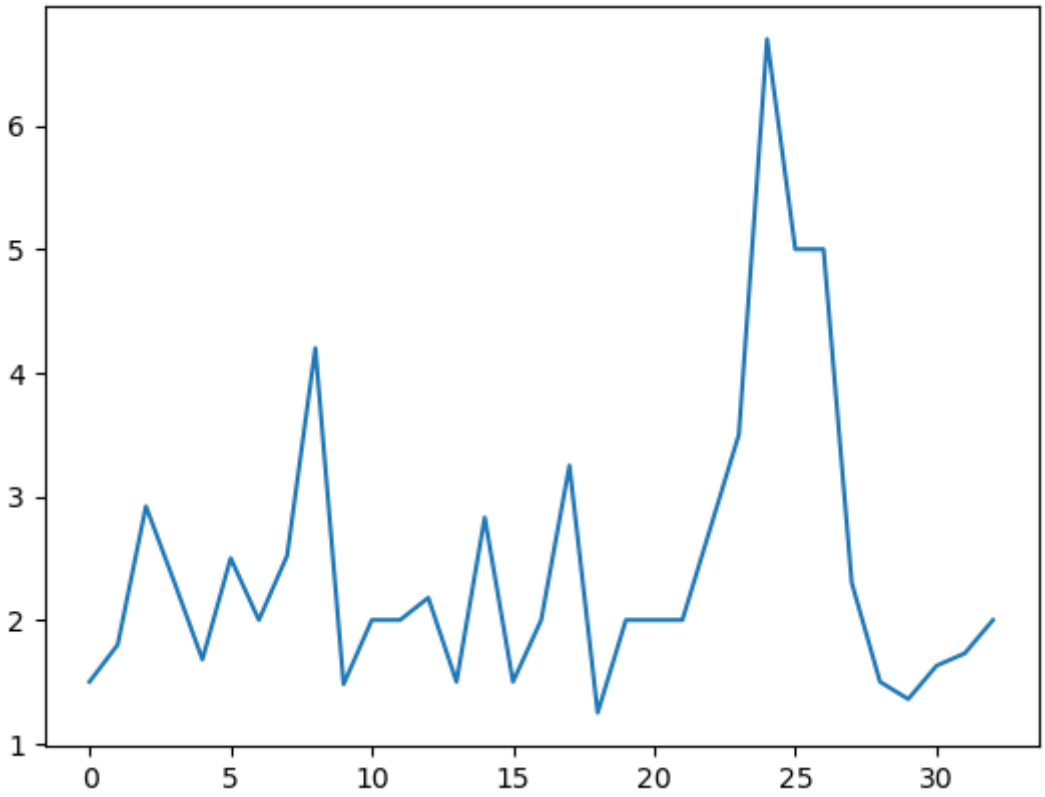

7
6
5
4
3
2
1
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

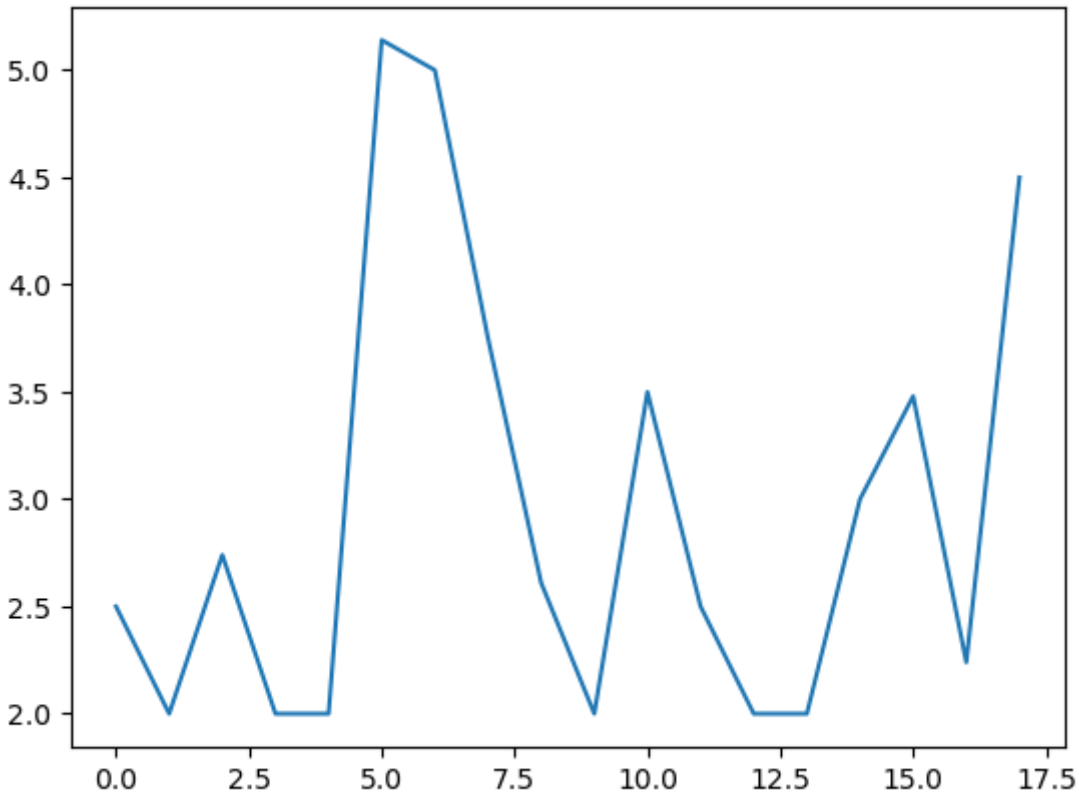

6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
0
2
4
6
8
10
12

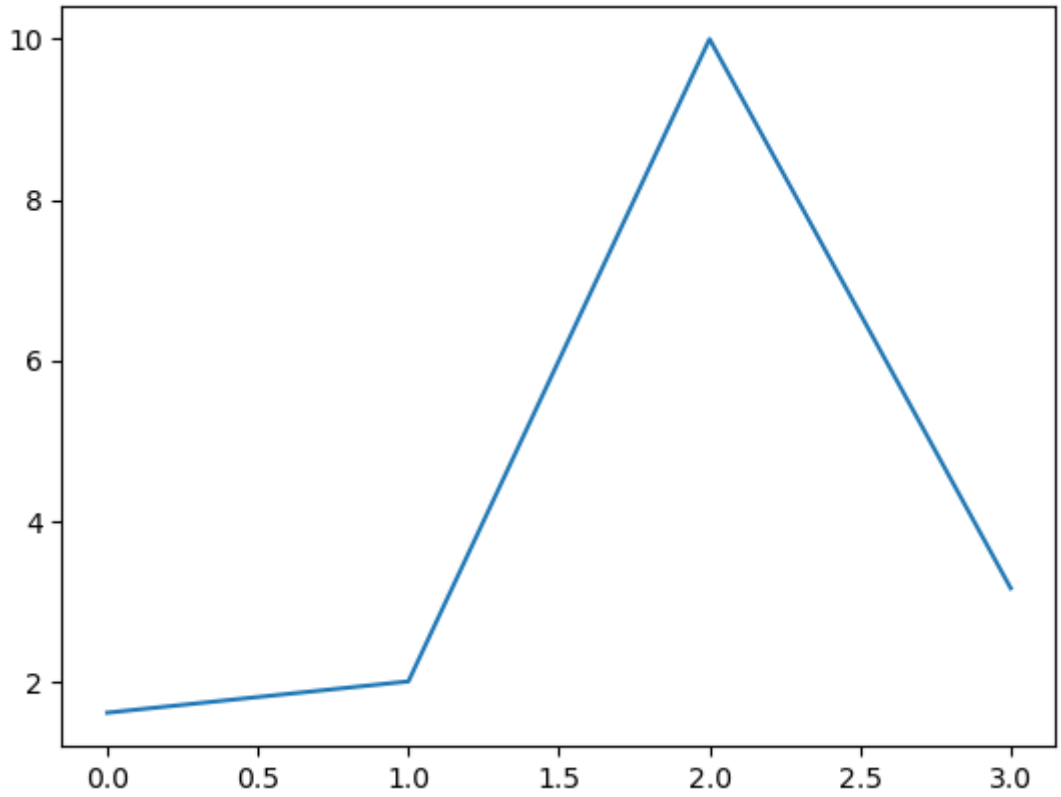

4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
2
4
6
8
10

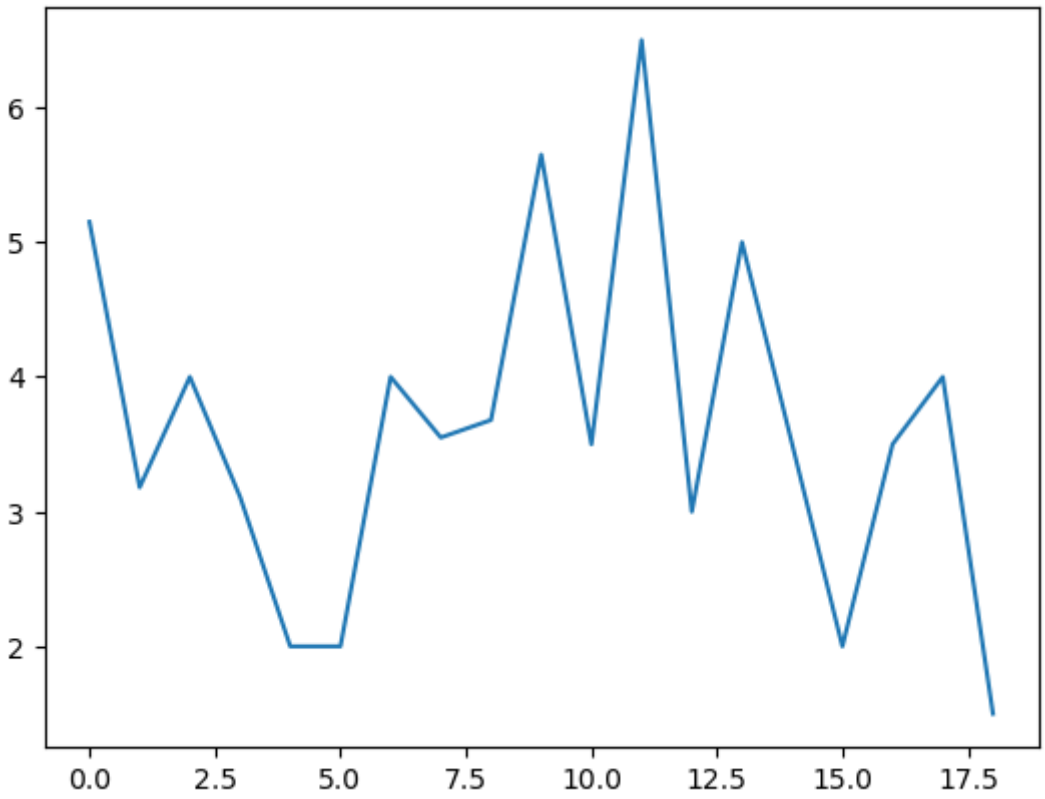

4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
2
4
6
8

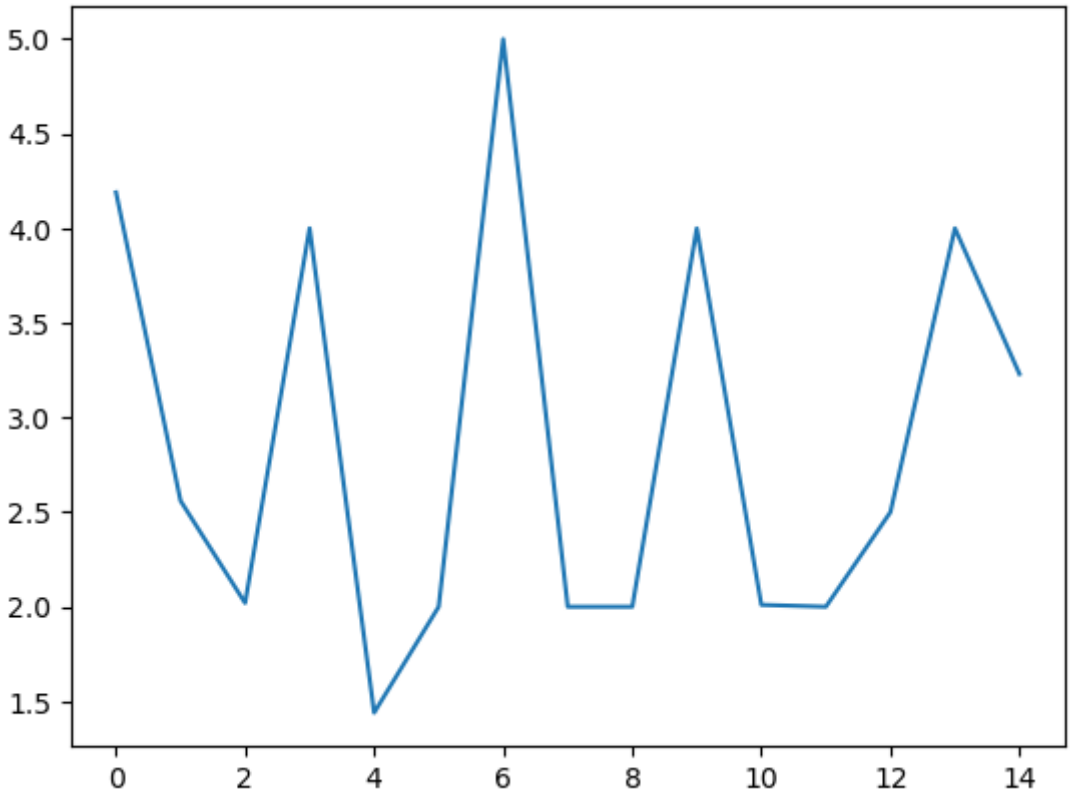

5.0
4.5
4.0
3.5
3.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

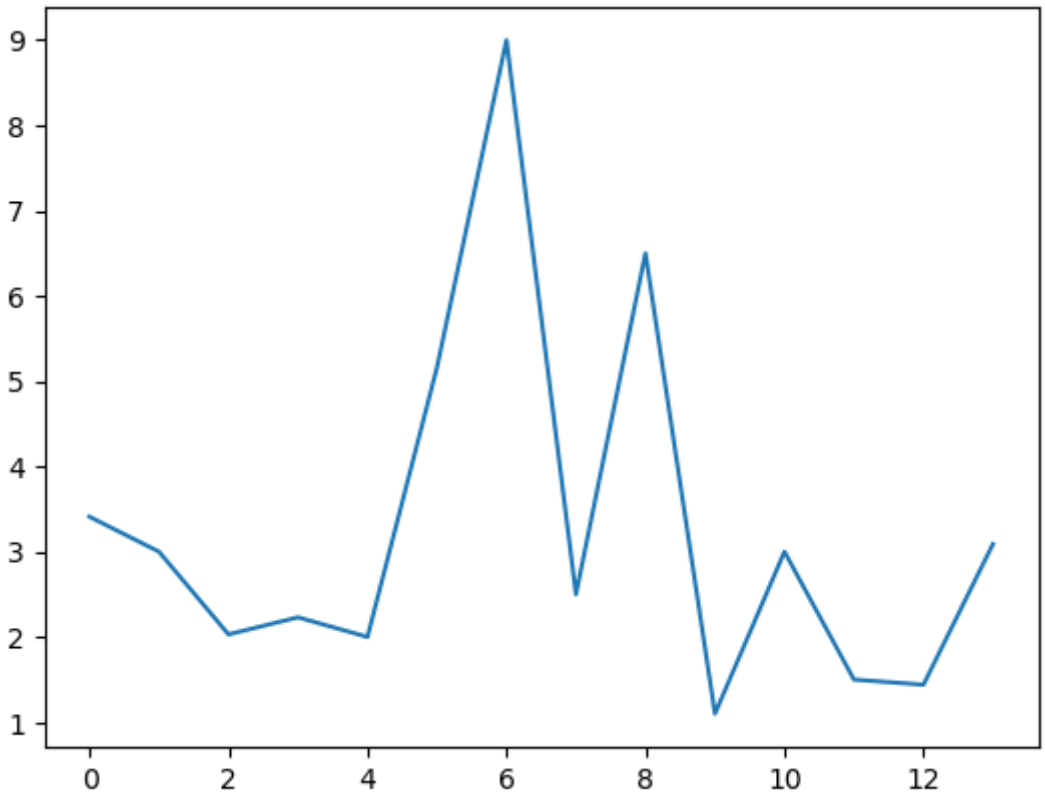

6
5
4
3
2
1
0
5
10
15
20
25
30

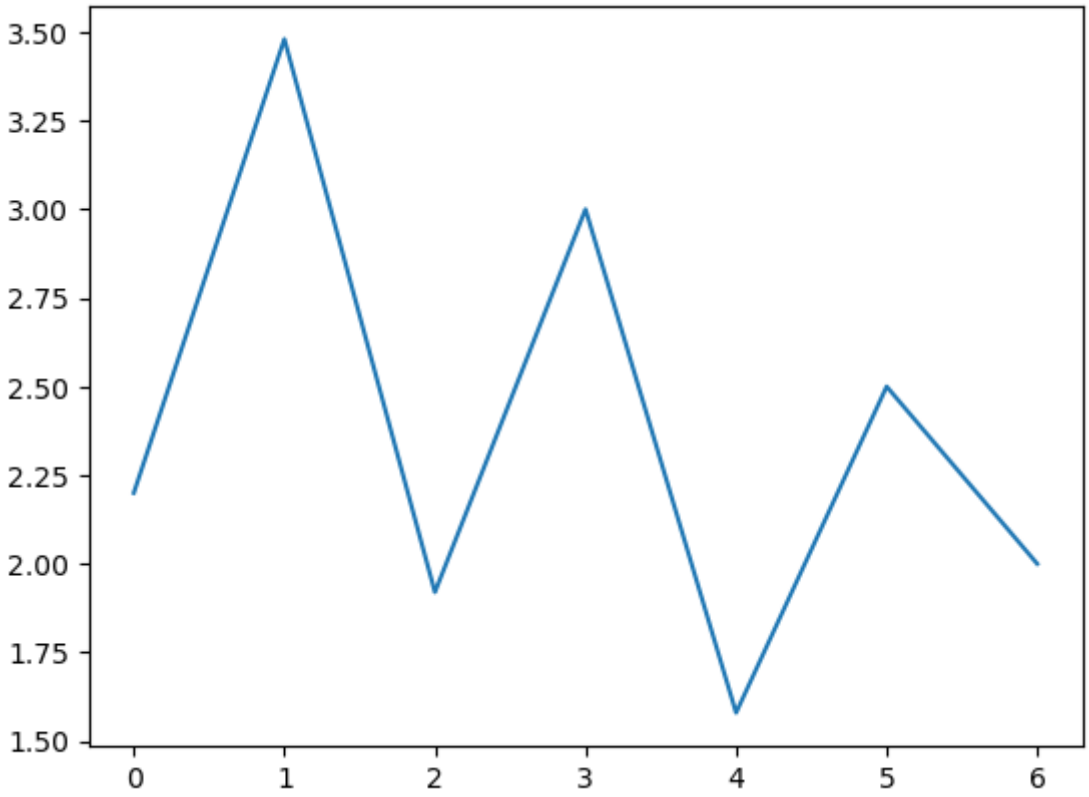

5.0
4.5
4.0
3.5
3.0
2.5
2.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5

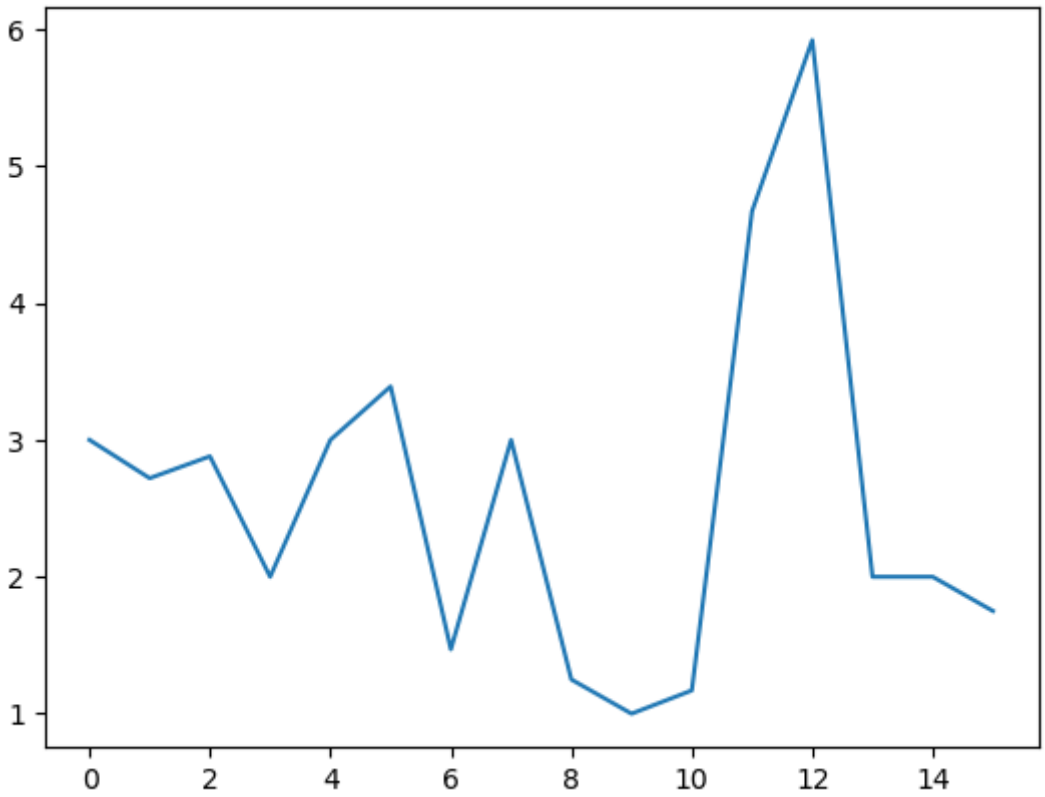

10
8
6
4
2
0.0
0.5
1.0
1.5
2.0
2.5
3.0


6
5
4
3
2
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5


5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0
2
4
6
8
10
12
14


9
8
7
6
5
4
3
2
1
0
2
4
6
8
10
12





we can see that tips are not increasing with time in each day, it's irregular

1. is there any difference in correlation between tip amounts from smokers and non-smokers?

In [157…
```
print("the correlation between meal prices and tips for smokers is",df[df['smok
print("the correlation between meal prices and tips for non-smokers is",df[df['
```

```
the correlation between meal prices and tips for smokers is 0.48821794116281
the correlation between meal prices and tips for non-smokers is 0.822182625705
0829
```