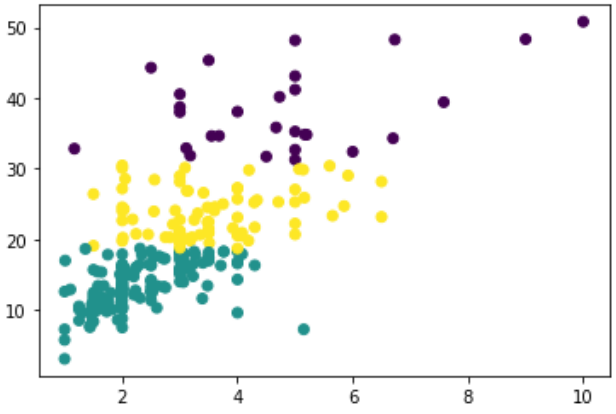
1. Using the dataset “tips.csv” implement k-means clustering technique on two fields “total_bill”, “tip”.

In [8]:
```
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
df = pd.read_csv(r'C:\Users\Nirmalya Majhi\Desktop\Advanced IT Workshop\tips.csv')
print(df.head())
x = 'tip'
y = 'total_bill'
kmeans = KMeans(init="random",n_clusters=3,n_init=10,max_iter=300,random_state=42)
kmeans.fit(df[[x,y]])
print(kmeans.labels_)
plt.scatter(df[x],df[y],c=kmeans.labels_)
plt.show()
```

```
   total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
[1 1 2 2 2 2 1 2 1 1 1 0 1 1 1 2 1 1 1 2 1 2 1 0 2 1 1 1 2 2 1 1 1 2 1 2 1
 1 1 0 1 1 1 1 2 1 2 0 2 1 1 1 0 1 2 2 0 2 1 0 2 1 1 1 1 2 1 1 2 1 1 1 2 2
 1 1 1 2 2 1 2 1 1 0 1 0 1 1 2 2 2 2 1 1 2 0 2 1 2 1 1 1 0 2 2 1 2 2 1 1 1
 1 0 2 2 1 2 1 1 2 1 1 1 1 1 2 1 1 1 2 2 2 1 1 1 1 1 1 1 1 0 0 2 1 1 1 1
 1 1 1 1 1 2 2 2 0 2 1 1 2 1 1 1 1 2 2 0 1 1 0 1 1 0 1 0 1 1 1 0 0 2 0 2 0
 2 2 2 1 2 1 2 2 1 1 1 1 0 1 1 2 1 1 1 2 1 2 0 2 1 2 2 0 1 2 1 2 1 1 2 1 1
 1 1 1 1 1 2 1 2 2 1 1 1 1 1 1 0 0 2 2 2 1 2]
```



In [ ]: