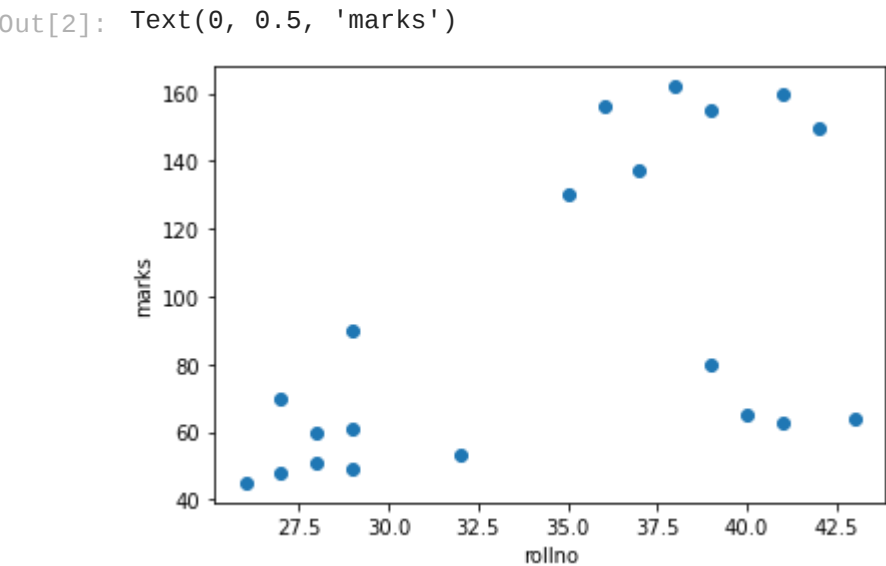


```
In [1]: from sklearn.cluster import KMeans
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
from matplotlib import pyplot as plt
df = pd.read_csv("Book1.csv")
df.head()
```

Out[1]:

	name	rollno	marks
0	A	40	65
1	B	41	63
2	C	43	64
3	D	39	80
4	E	36	156

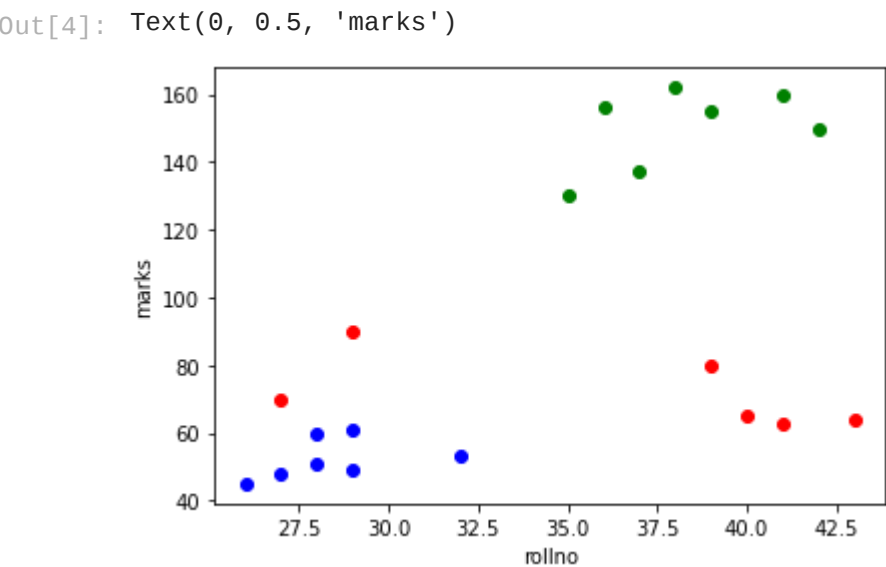
```
In [2]: plt.scatter(df.rollno,df['marks'])
plt.xlabel('rollno')
plt.ylabel('marks')
```



```
In [3]: km = KMeans(n_clusters=3)
predicted = km.fit_predict(df[['rollno', 'marks']])
predicted
```

Out[3]: array([1, 1, 1, 1, 0, 0, 0, 2, 2, 2, 2, 2, 1, 1, 2, 2, 0, 0, 0, 0])

```
In [4]: df['cluster']=predicted
df.head()
df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.rollno,df1['marks'],color='green')
plt.scatter(df2.rollno,df2['marks'],color='red')
plt.scatter(df3.rollno,df3['marks'],color='blue')
plt.xlabel('rollno')
plt.ylabel('marks')
```



```
In [5]: scale = MinMaxScaler()

scale.fit(df[['marks']])
df['marks'] = scale.transform(df[['marks']])

scale.fit(df[['rollno']])
df['rollno'] = scale.transform(df[['rollno']])
```

```
In [6]: km = KMeans(n_clusters=3)
predicted = km.fit_predict(df[['rollno', 'marks']])
predicted
```

Out[6]: array([2, 2, 2, 2, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1])

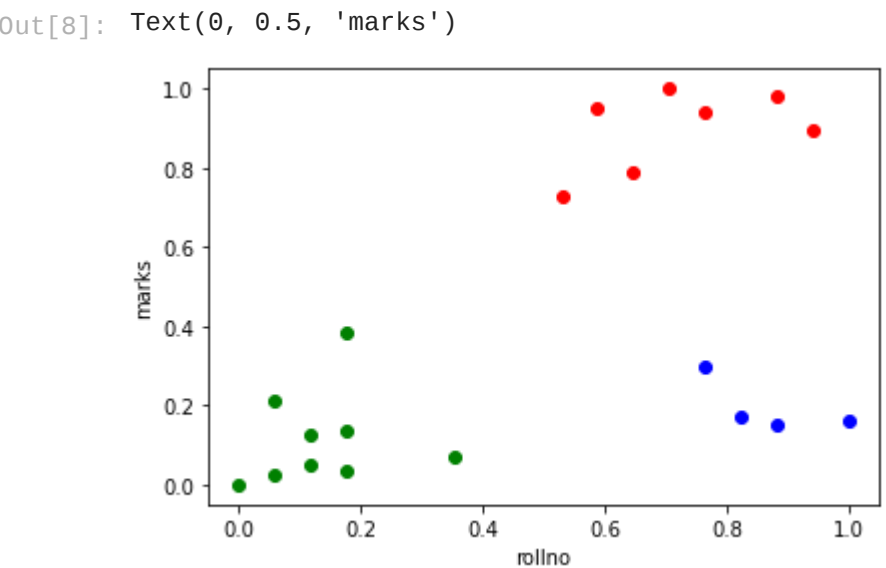
```
In [7]: df = df.drop(['cluster'], axis='columns')

df['cluster']=predicted
df.head()
```

Out[7]:

	name	rollno	marks	cluster
0	A	0.823529	0.170940	2
1	B	0.882353	0.153846	2
2	C	1.000000	0.162393	2
3	D	0.764706	0.299145	2
4	E	0.588235	0.948718	1

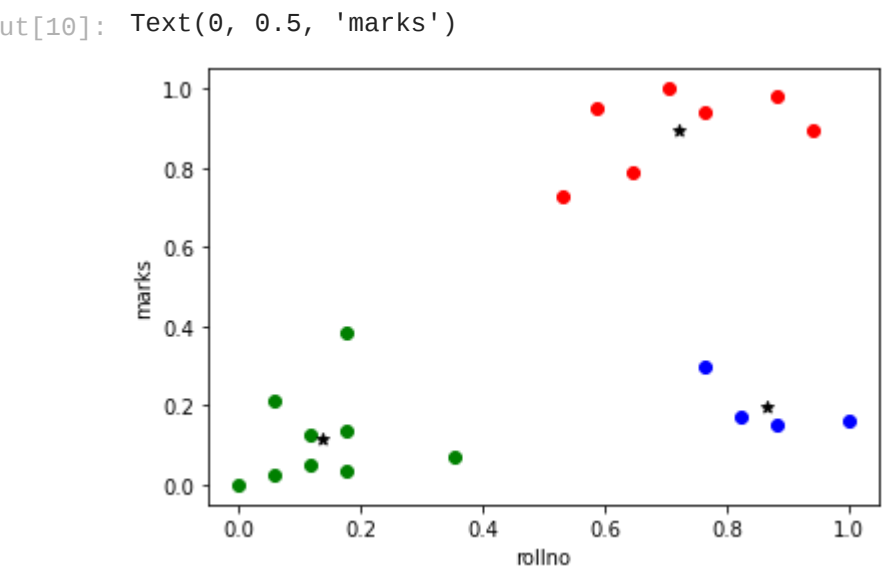
```
In [8]: df1 = df[df.cluster==0]
df2 = df[df.cluster==1]
df3 = df[df.cluster==2]
plt.scatter(df1.rollno,df1['marks'],color='green')
plt.scatter(df2.rollno,df2['marks'],color='red')
plt.scatter(df3.rollno,df3['marks'],color='blue')
plt.xlabel('rollno')
plt.ylabel('marks')
```



```
In [9]: km.cluster_centers_
```

Out[9]: array([[0.1372549 , 0.11585945],
[0.72268908, 0.8974359],
[0.86764706, 0.1965812]])

```
In [10]: plt.scatter(df1.rollno,df1['marks'],color='green')
plt.scatter(df2.rollno,df2['marks'],color='red')
plt.scatter(df3.rollno,df3['marks'],color='blue')
plt.scatter(km.cluster_centers_[0],km.cluster_centers_[1],color='black',marker='*')
plt.xlabel('rollno')
plt.ylabel('marks')
```



```
In [ ]:
```