

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

ds=pd.read_csv('Kmeanscluster.csv')
print(ds.shape)
print(ds.head(5))
print(ds.describe())
```

```
(20, 2)
   Income  Spend
0      285    170
1      260    163
2      270    156
3      245    165
4      250    155
```

	Income	Spend
count	20.000000	20.000000
mean	255.700000	159.600000
std	27.158599	19.427057
min	198.000000	123.000000
25%	233.750000	151.750000
50%	260.000000	156.000000
75%	273.750000	172.250000
max	296.000000	195.000000

```
Income=ds['Income'].values
Spend=ds['Spend'].values
x=np.array(list(zip(Income,Spend)))
x
```

```
array([[285, 170],
       [260, 163],
       [270, 156],
       [245, 165],
       [250, 155],
       [230, 130],
       [280, 129],
       [295, 180],
       [225, 195],
       [260, 171],
       [267, 156],
       [259, 123],
       [233, 154],
       [276, 176],
       [296, 190],
       [210, 178],
       [198, 145],
       [268, 155],
       [273, 156],
       [234, 145]])
```

```
from sklearn.cluster import KMeans
wcss=[]
for i in range(1,11):
    Km=KMeans(n_clusters=1,random_state=0)
    Km.fit(x)
    wcss.append(Km.inertia_)
plt.plot(range(1,11),wcss,color='green',marker='.')
plt.xlabel('Number of clusters')
plt.ylabel('wcss')
plt.title('Optimal K values')
plt.show()
```



Optimal K values

22000

```
model=KMeans(n_clusters=5,random_state=0)
y_means=model.fit_predict(x)
```

```
plt.scatter(x[y_means==0],x[y_means==0,1],s=50,c='violet',label='cluster1')
plt.scatter(x[y_means==1],x[y_means==1,1],s=50,c='green',label='cluster2')
plt.scatter(x[y_means==2],x[y_means==2,1],s=50,c='pink',label='cluster3')
plt.scatter(x[y_means==3],x[y_means==3,1],s=50,c='orange',label='cluster4')
plt.title('Income spend Analysis')
plt.xlabel('Income')
plt.ylabel('Spend')
plt.legend()
plt.show()
```

