```
In [1]:
```

```python
import numpy as np
import pandas as pd
```

```
In [2]:
```

```python
df = pd.read_csv('Mall_Customers.csv')
df.shape
```

```
Out[2]:
```

```
(200, 5)
```

```
In [3]:
```

```python
df.head()
```

```
Out[3]:
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

```
In [4]:
```

```python
df["A"]= df[["Annual Income (k$)"]]
df["B"]=df[["Spending Score (1-100)"]]
```

```
In [5]:
```

```python
X=df[["A","B"]]
X.head()
```
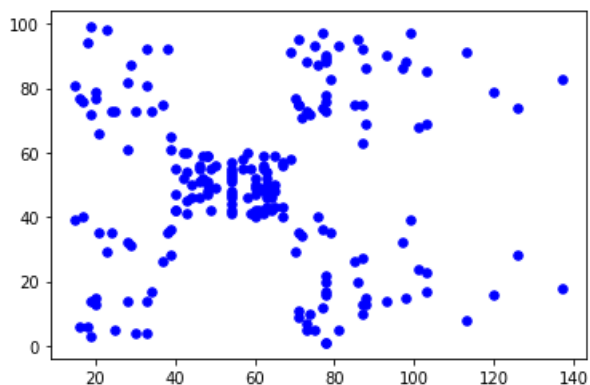
```
Out[5]:
```

	A	B
0	15	39
1	15	81
2	16	6
3	16	77
4	17	40

```
In [6]:
```

```python
# Commented out IPython magic to ensure Python compatibility.
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
# %matplotlib inline
```

```
In [7]:
```

```python
plt.scatter(X["A"], X["B"], s = 30, c = 'b')
plt.show()
```



```
In [8]:
```

```
Kmean = KMeans(n_clusters=5)
Kmean.fit(X)
```

Out[8]:

```
KMeans(n_clusters=5)
```

In [9]:

```
centers=Kmean.cluster_centers_
print(Kmean.cluster_centers_)
```

```
[[88.2        17.11428571]
 [55.2962963  49.51851852]
 [26.30434783 20.91304348]
 [86.53846154 82.12820513]
 [25.72727273 79.36363636]]
```

In [10]:

```
clusters = Kmean.fit_predict(X)
df["label"] = clusters
df.head(100)
```

Out[10]:

|	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)	A	B	label
0	1	Male	19	15	39	15	39	3
1	2	Male	21	15	81	15	81	4
2	3	Female	20	16	6	16	6	3
3	4	Female	23	16	77	16	77	4
4	5	Female	31	17	40	17	40	3
...	...	...	...	...	...	...	...	...
95	96	Male	24	60	52	60	52	0
96	97	Female	47	60	47	60	47	0
97	98	Female	27	60	50	60	50	0
98	99	Male	48	61	42	61	42	0
99	100	Male	20	61	49	61	49	0

100 rows × 8 columns

In [11]:

```
col=['green','blue','black','yellow','orange',]
```

In [12]:

```
for i in range(5):
    a=col[i]
    # print(a)
    plt.scatter(df.A[df.label==i], df.B[df.label == i], c=a, label='cluster 1')
plt.scatter(centers[:, 0], centers[:, 1], marker='*', s=300, c='r', label='centroid')
```

Out[12]:

```
<matplotlib.collections.PathCollection at 0x7f47c1c47910>
```
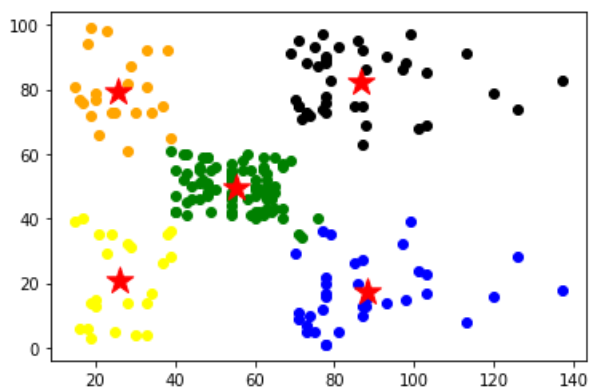

In [13]:

```
X1 = X.loc[:,["A","B"]].values
```

In [14]:

```python
wcss=[]
for k in range(1,11):
    kmeans = KMeans(n_clusters = k, init = "k-means++")
    kmeans.fit(X1)
    wcss.append(kmeans.inertia_)
plt.figure(figsize =( 12,6))
plt.grid()
plt.plot(range(1,11),wcss,linewidth=2,color="red",marker="8")
plt.xlabel("K Value")
plt.ylabel("WCSS")
plt.show()
```