

kmeans_sales.py

```
# -*- coding: utf-8 -*-  
"""kmeans_sales.ipynb
```

Automatically generated by Colaboratory.

Original file is located at
<https://colab.research.google.com/drive/1Ibt53yv3fuLNRXwpjXLLhFrdyBSGIN6P>
"""

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

```
from google.colab import files  
uploaded = files.upload()
```

```
import io  
df = pd.read_csv(io.BytesIO(uploaded['Mall_Customers.csv']))  
df.shape
```

```
df.head()
```

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

```
X = df[["A", "B"]]  
X.head()
```

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

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

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

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

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

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

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