

## Assignment 4

```
# Clustering Assignment - Single Script Version
# Converted from Jupyter Notebook to Python script
# -----

import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)
warnings.filterwarnings('ignore')

df = pd.read_csv(r"C:\Users\rohit\OneDrive\Desktop\Sem 5\3.End_Sem
ML\Mall_Customers.csv")

df

x = df.iloc[:,3:]

x

plt.title('Unclustered Data')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.scatter(x['Annual Income (k$)'], x['Spending Score (1-100)'])

from sklearn.cluster import KMeans, AgglomerativeClustering
km = KMeans(n_clusters=6)
km.fit_predict(x)

#SSE
km.inertia_

sse = []
for k in range(1, 16):
    km = KMeans(n_clusters=k)
    km.fit_predict(x)
    sse.append(km.inertia_)

sse
```

```

plt.title('Elbow Method')
plt.xlabel('Value of K')
plt.ylabel('SSE')
plt.grid()
plt.xticks(range(1,16))
plt.plot(range(1,16), sse, marker='.', color='red')

```

```

from sklearn.metrics import silhouette_score

```

```

silh = []
for k in range(2, 16):
    km = KMeans(n_clusters=k)
    labels = km.fit_predict(x)
    score = silhouette_score(x, labels)
    silh.append(score)

```

```

plt.title('Silhouette Method')
plt.xlabel('value of K')
plt.ylabel('Silhouette Score')
plt.xticks(range(2, 16))
plt.plot(range(2, 16), silh, color='red')

```

```

km = KMeans(n_clusters=5, random_state=0)

```

```

labels = km.fit_predict(x)

```

```

labels

```

```

km.cluster_centers_

```

```

cent = km.cluster_centers_

```

```

plt.figure(figsize=(16, 9))
plt.subplot(1, 2, 1)
plt.title('Unclustered Data')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.scatter(x['Annual Income (k$)'], x['Spending Score (1-100)'])
plt.subplot(1, 2, 2)
plt.title('Clustered Data')
plt.xlabel('Annual Income')
plt.ylabel('Spending Score')
plt.scatter(x['Annual Income (k$)'], x['Spending Score (1-100)'], c=labels)

```

```
plt.scatter(cent[:,0], cent[:,1], s = 100, c='crimson')
```

```
km.inertia_
```

```
km.labels_
```

```
four = df[labels==4]  
four.to_csv('mydata.csv')  
km.predict([[46, 78]])
```

```
agl = AgglomerativeClustering(n_clusters=5)
```

```
alabels = agl.fit_predict(x)
```

```
alabels
```

```
plt.figure(figsize=(16, 9))
```

```
plt.subplot(1, 2, 1)  
plt.title('Agglomerative')  
plt.xlabel('Annual Income')  
plt.ylabel('Spending Score')  
plt.scatter(x['Annual Income (k$)'], x['Spending Score (1-100)'], c=alabels)  
plt.subplot(1, 2, 2)  
plt.title('KMeans')  
plt.xlabel('Annual Income')  
plt.ylabel('Spending Score')  
plt.scatter(x['Annual Income (k$)'], x['Spending Score (1-100)'], c=labels)  
plt.scatter(cent[:,0], cent[:,1], s = 100, c='crimson')
```