21162101012_CBA_Yash_Lakhtariya

DMW Practical 9

A dataset of a Mall (File Name-"Mall_Customer.csv") is provided to you for this practical. It contains the information about the customer with different attributes such as customer ID, Age, Gender, annual income and as per the spending the customer has done, some spending scores are assigned to them. As a data analyst, you need to find out similar kind of customers. This information will be helpful to target a specific group of customers for the sale of products.But the question here is how the customer will be identified and grouped? Which attributes is contributing the most to identify the group of customers? how many such groups can be created? Try to answer the questions and create the group of customers as per the details provided in the dataset. Also, provide the visualization of the groups that are being made.

```
[ ] from google.colab import files
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    import seaborn as sns
    import io
```

```
[ ] uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

```
Saving Mall_Customers.csv to Mall_Customers.csv
```

```
[ ] dataset = pd.read_csv(io.BytesIO(uploaded['Mall_Customers.csv']))
```

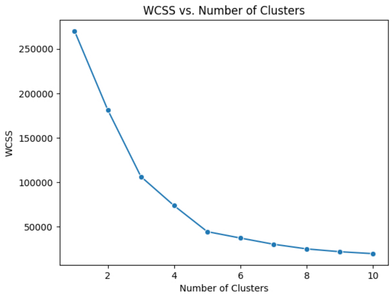
```
[ ] data_matrix = dataset.iloc[:, [3, 4]].values
    data_matrix
```

```
array([[ 15,  39],
       [ 15,  81],
       [ 16,   6],
       [ 16,  77],
       [ 17,  40],
       [ 17,  76],
       [ 18,   6],
       [ 18,  94],
       [ 19,   3],
       [ 19,  72],
       [ 19,  14],
       [ 19,  99],
       [ 20,  15],
       [ 20,  77],
       [ 20,  13],
       [ 20,  79],
       [ 21,  35],
       [ 21,  66],
       [ 23,  29],
       [ 23,  98],
       [ 24,  35],
       [ 24,  73],
       [ 25,   5],
       [ 25,  73],
       [ 28,  14],
       [ 28,  82],
       [ 28,  32],
       [ 28,  61],
       [ 29,  31],
       [ 29,  87],
       [ 30,   4],
       [ 30,  73],
       [ 33,   4],
       [ 33,  92],
       [ 33,  14],
       [ 33,  81],
       [ 34,  17],
       [ 34,  73],
       [ 37,  26],
       [ 37,  75],
       [ 38,  35],
       [ 38,  92],
       [ 39,  36],
       [ 39,  61],
       [ 39,  28],
       [ 39,  65],
       [ 40,  55],
       [ 40,  47],
       [ 40,  42],
       [ 40,  42],
       [ 42,  52],
       [ 42,  60],
       [ 43,  54],
       [ 43,  60],
       [ 43,  45],
       [ 43,  41],
       [ 44,  50],
       [ 44,  46],
```

```
[ ] from sklearn.cluster import KMeans
```

```
[ ] wcss = []
    for i in range(1, 11):
        kmeans = KMeans(n_clusters=i, random_state=42)
        kmeans.fit(data_matrix)
        wcss.append(kmeans.inertia_)
```

```
[ ] sns.lineplot(x=range(1, 11), y=wcss, marker='o')
    plt.title("WCSS vs. Number of Clusters")
    plt.xlabel("Number of Clusters")
    plt.ylabel("WCSS")
    plt.show()
```




```
[ ] kmeans = KMeans(n_clusters=5, random_state=42)
    y_kmeans = kmeans.fit_predict(data_matrix)
```

```
[ ] colors = ['blue', 'green', 'yellow', 'red', 'pink']
    for i, color in enumerate(colors):
        plt.scatter(data_matrix[y_kmeans == i, 0], data_matrix[y_kmeans == i, 1], s=100, c=color, label=f'Cluster {i+1}')

    plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s=200, c='black', marker='s', label='Centroids')
    plt.title('Cluster of Customers')
    plt.xlabel('Annual Income (k$)')
    plt.ylabel('Spending Score')
    plt.legend()
    plt.show()
```