Name : Gayatri Bhagwat

Week 2 Task

CUSTOMER SEGMENTATION USING CLUSTERING

Description: Use K–Means or hierarchical clustering to group customers based on purchasing behavior. **Challenges**: Scale and normalize features Choose optimal k using the Elbow Method Visualize clusters using 2D/3D plots Tech Stack: Python, Scikit-learn, Pandas, Matplotlib

```python
# Importing Liabraries
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

```python
# Loding the Dataset
df = pd.read_csv("/content/Mall_Customers.csv")
df.head()
```

	CustomerID	Gender	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

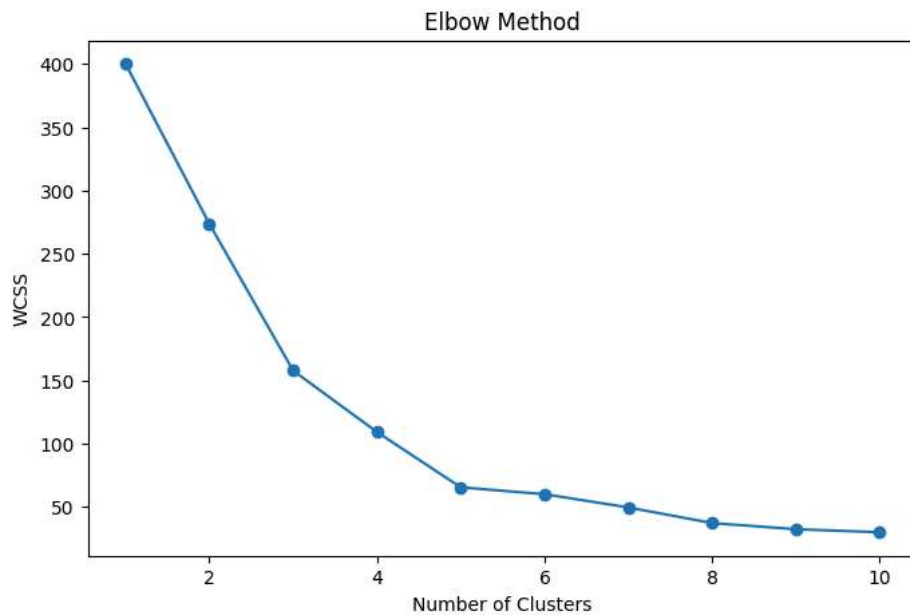Next steps: (Generate code with df) (New interactive sheet)

```python
# Select Features from Clustering
X = df[['Annual Income (k$)', 'Spending Score (1-100)']]
```

```python
# Feature Scaling
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

```python
# Elbow Method
wcss = []

for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

plt.figure(figsize=(8,5))
plt.plot(range(1,11), wcss, marker='o')
plt.xlabel('Number of Clusters')
plt.ylabel('WCSS')
plt.title('Elbow Method')
plt.show()
```
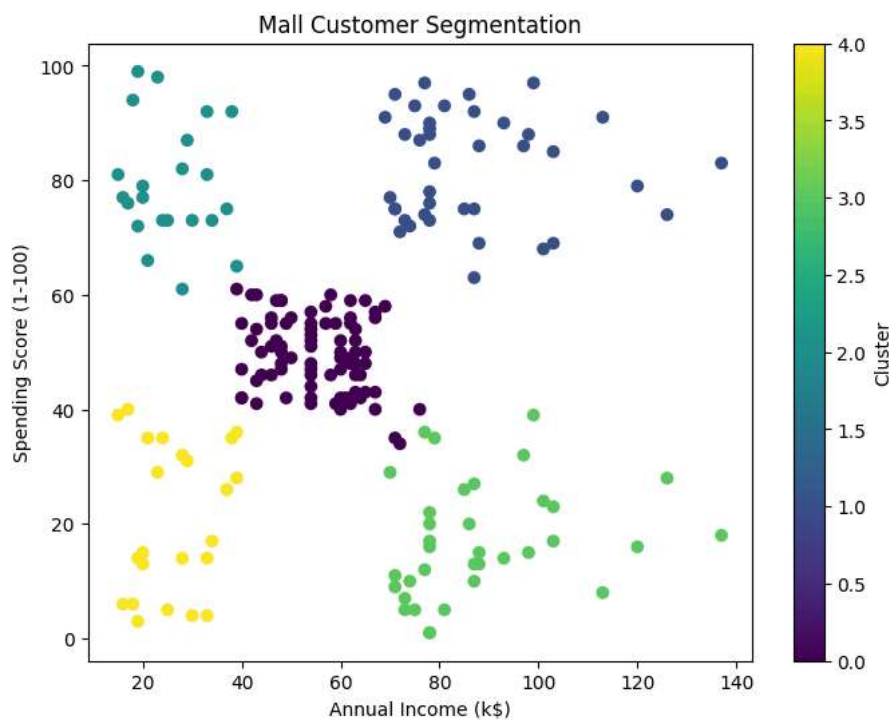
Elbow Method

```python
# k-means Clustering
kmeans = KMeans(n_clusters=5, random_state=42)
df['Cluster'] = kmeans.fit_predict(X_scaled)
```

```python
#Visualize Customer Segments (2D)
plt.figure(figsize=(8,6))
plt.scatter(
    df['Annual Income (k$)'],
    df['Spending Score (1-100)'],
    c=df['Cluster'],
    cmap='viridis'
)
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.title('Mall Customer Segmentation')
plt.colorbar(label='Cluster')
plt.show()
```


Mall Customer Segmentation

```
#Cluster Interpretation
df.groupby('Cluster')[['Annual Income (k$)', 'Spending Score (1-100)']].mean()
```

	Annual Income (k$)	Spending Score (1-100)	
Cluster			
0	55.296296	49.518519	
1	86.538462	82.128205	
2	25.727273	79.363636	
3	88.200000	17.114286	
4	26.304348	20.913043	

```python
fig = plt.figure(figsize=(10,8))
ax = fig.add_subplot(111, projection='3d')

ax.scatter(
    df['Age'],
    df['Annual Income (k$)'],
    df['Spending Score (1-100)'],
    c=df['Cluster'],
    cmap='rainbow'
)

ax.set_xlabel('Age')
ax.set_ylabel('Annual Income (k$)')

# RIGHT-SIDE Z LABEL (manual & clear)
fig.text(0.84, 0.5, 'Spending Score (1-100)',
         va='center', rotation='vertical')

plt.title('3D Mall Customer Segmentation')
plt.show()
```


3D Mall Customer Segmentation