Assignment No. : 6

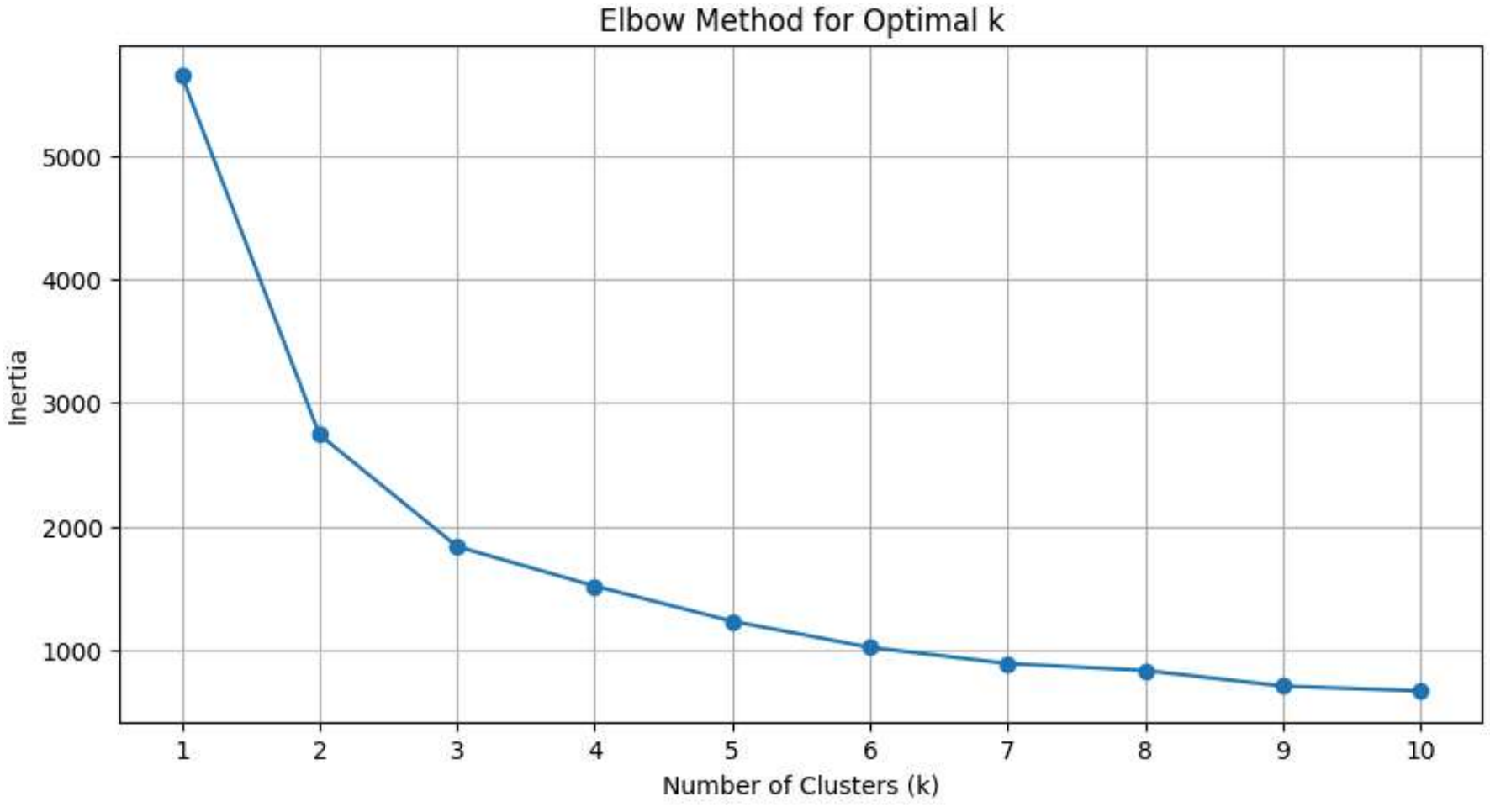
Implement K-Means clustering/ hierarchical clustering on sales_data_sample.csv dataset. Determine the number of clusters using the elbow method. Dataset link : https://www.kaggle.com/datasets/kyanyoga/sample-sales-data

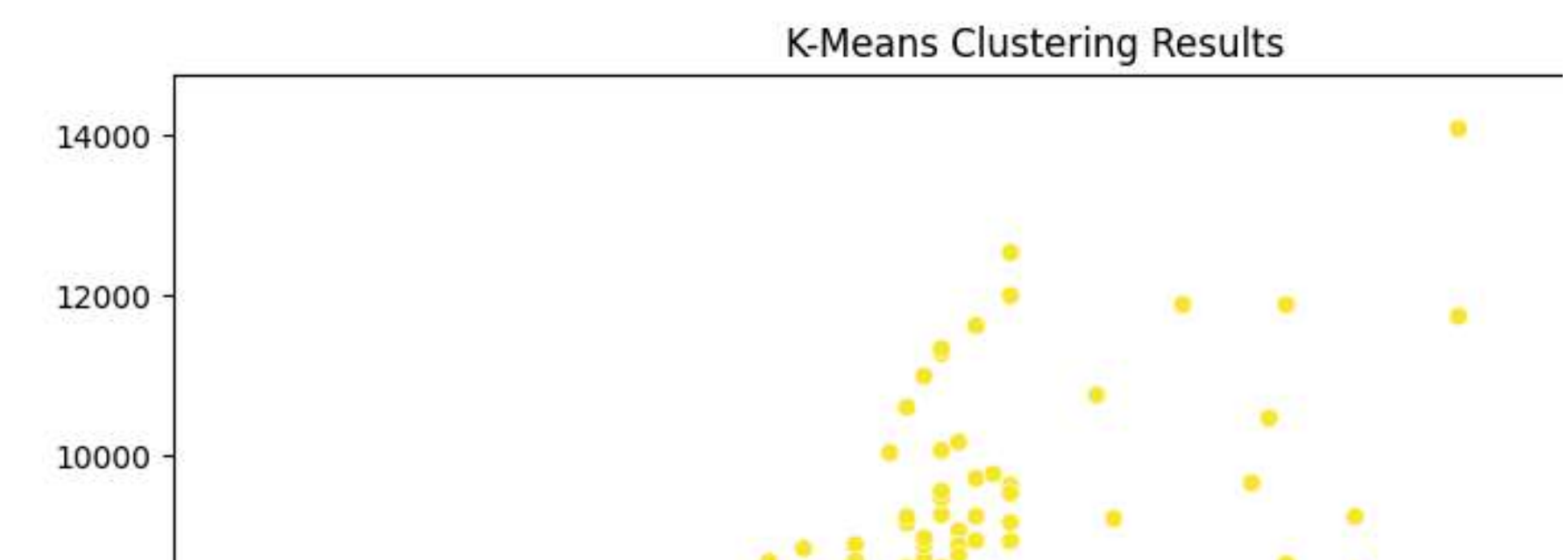
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
# Step 1: Load the Dataset
data = pd.read_csv("/content/sales_data_sample.csv", encoding='latin1')
data.head(2)
```

```
# Drop rows with NaN values
# Step 2: Preprocess the Data
# Select relevant features for clustering
features = data[['QUANTITYORDERED', 'SALES']].dropna()
# Normalize the data
scaler = StandardScaler()
scaled_features = scaler.fit_transform(features)
```

```
# Step 3: Elbow Method to Determine Optimal Number of Clusters
inertia = []
K = range(1, 11)
for k in K:
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(scaled_features)
    inertia.append(kmeans.inertia_)
    # Plot the elbow graph
plt.figure(figsize=(10, 5))
plt.plot(K, inertia, marker='o')
plt.title('Elbow Method for Optimal k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Inertia')
plt.xticks(K)
plt.grid()
plt.show()

# Step 4: Perform K-Means Clustering with the Optimal Number of Clusters
optimal_k = 3  # Set this based on the elbow method result
kmeans = KMeans(n_clusters=optimal_k, random_state=42)
clusters = kmeans.fit_predict(scaled_features)
# Add the cluster labels to the original dataset
data['Cluster'] = pd.Series(clusters)
# Step 5: Visualize the Clusters
plt.figure(figsize=(10, 7))
sns.scatterplot(data=data, x='QUANTITYORDERED', y='SALES', hue='Cluster', palette='vi
plt.title('K-Means Clustering Results')
plt.xlabel('QUANTITY')
plt.ylabel('SALES')
plt.legend(title='Cluster')
plt.show()
```


Elbow Method for Optimal k
Inertia
5000
4000
3000
2000
1000
1
2
3
4
5
6
7
8
9
10
Number of Clusters (k)


K-Means Clustering Results
SALES
14000
12000
10000
8000
6000
4000
2000
0
20
40
60
80
QUANTITY