Code & OUTPUT

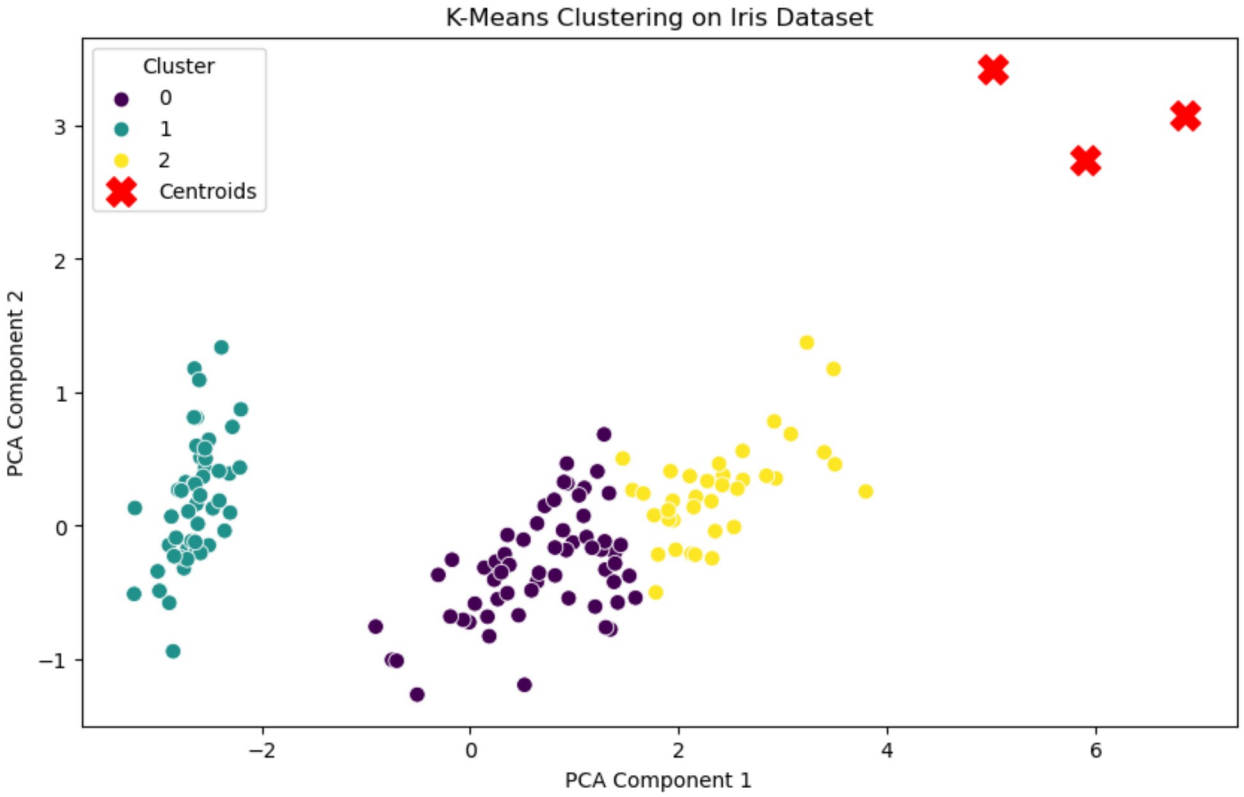
In [1]:
```
print("Experiment No 04 : To implement k means clustering.")
```

```
Experiment No 04 : To implement k means clustering.
```

In [5]:
```
# Import necessary libraries
import pandas as pd
from sklearn.datasets import load_iris
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.decomposition import PCA

print("OUTPUT:\n\n")

# Load the Iris dataset
iris = load_iris()
X = iris.data  # Features
y = iris.target  # True labels for comparison

# Initialize the KMeans model
# Assuming 3 clusters as there are 3 species in the Iris dataset
kmeans = KMeans(n_clusters=3, random_state=42)

# Fit the model and predict clusters
y_kmeans = kmeans.fit_predict(X)

# Show cluster centers
print("Cluster Centers:")
print(kmeans.cluster_centers_)

# Visualize the clusters in a 2D plot using PCA (Principal Component Analysis) for dimensionality reduction
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

# Plot the results
plt.figure(figsize=(10, 6))
sns.scatterplot(x=X_pca[:, 0], y=X_pca[:, 1], hue=y_kmeans, palette="viridis", s=60)
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], c='red', marker='X', s=200, label='Centroid
plt.title("K-Means Clustering on Iris Dataset")
plt.xlabel("PCA Component 1")
plt.ylabel("PCA Component 2")
plt.legend(title="Cluster")
plt.show()

# Evaluate by comparing with true labels
# Note: K-Means does not use true labels, this is just for evaluation purposes.
from sklearn.metrics import accuracy_score, confusion_matrix

# Map clusters to actual labels (for evaluation purposes only)
# In this case, we need to manually assign the clusters to labels for evaluation
# This can vary each time you run it
label_map = {0: 'setosa', 1: 'versicolor', 2: 'virginica'}
predicted_labels = [label_map[label] for label in y_kmeans]

# Display confusion matrix
print("\nConfusion Matrix:")
print(confusion_matrix(y, y_kmeans))
```

OUTPUT:

```
c:\Users\hp\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1412: FutureWarning: The default value of `n_init
` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
c:\Users\hp\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:1436: UserWarning: KMeans is known to have a memo
ry leak on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the envi
ronment variable OMP_NUM_THREADS=1.
  warnings.warn(
Cluster Centers:
[[5.9016129  2.7483871  4.39354839 1.43387097]
 [5.006      3.428      1.462      0.246     ]
 [6.85       3.07368421 5.74210526 2.07105263]]
```




```
Confusion Matrix:
[[ 0 50  0]
 [48  0  2]
 [14  0 36]]
```

In [ ]: