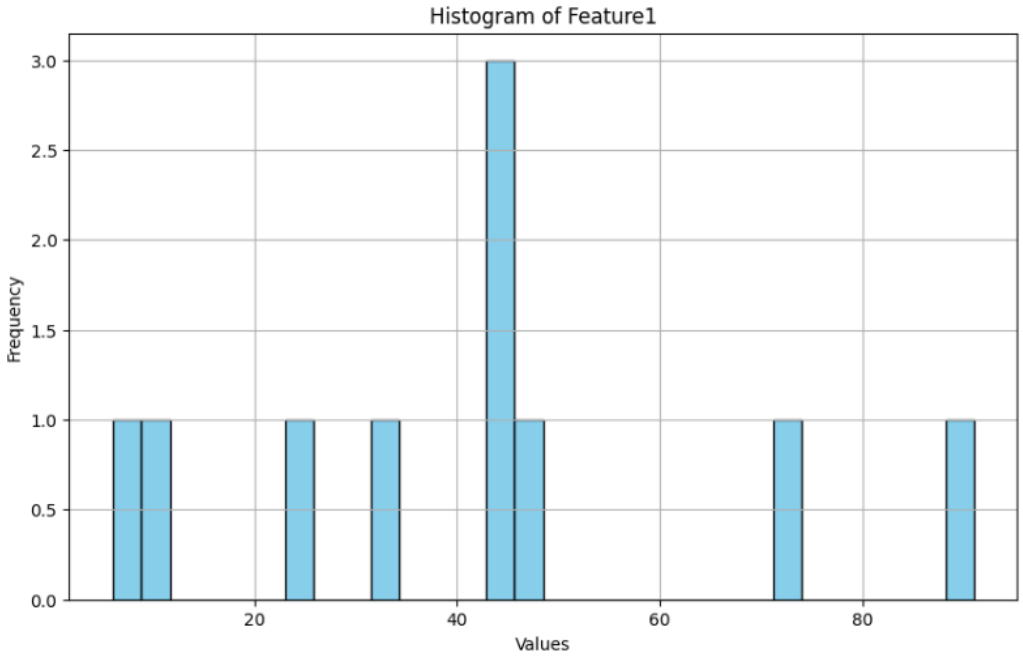
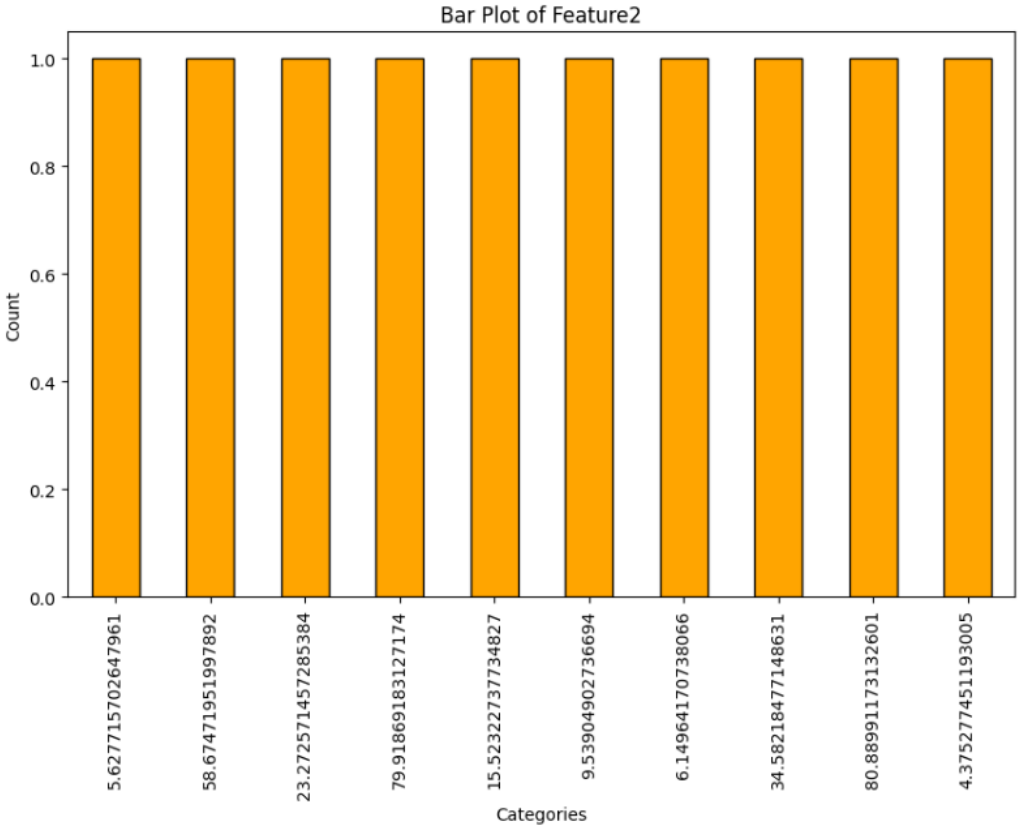
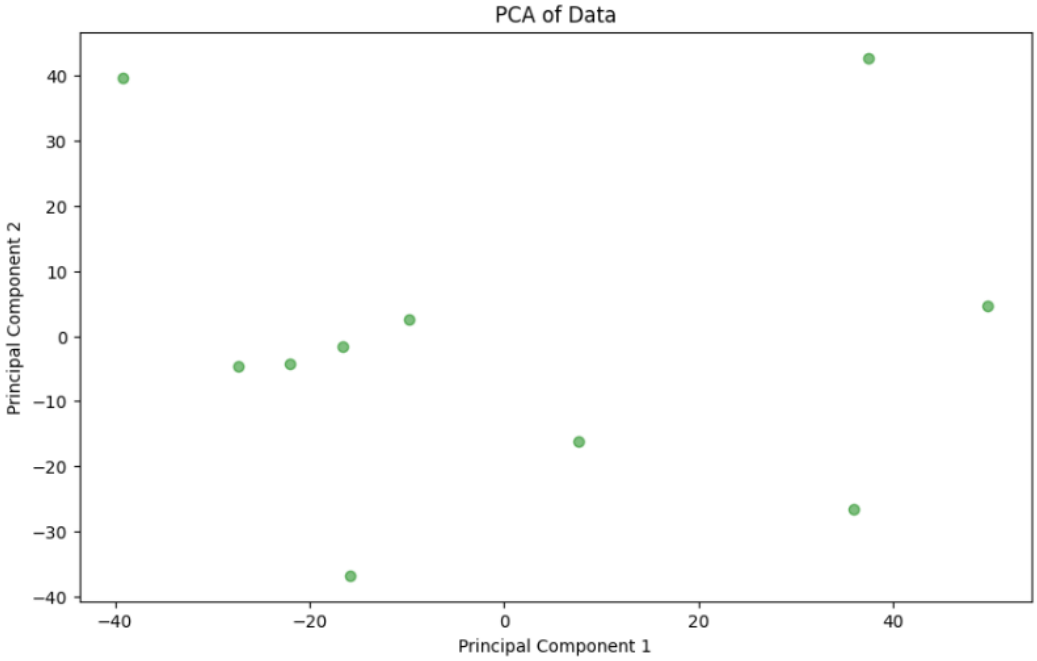
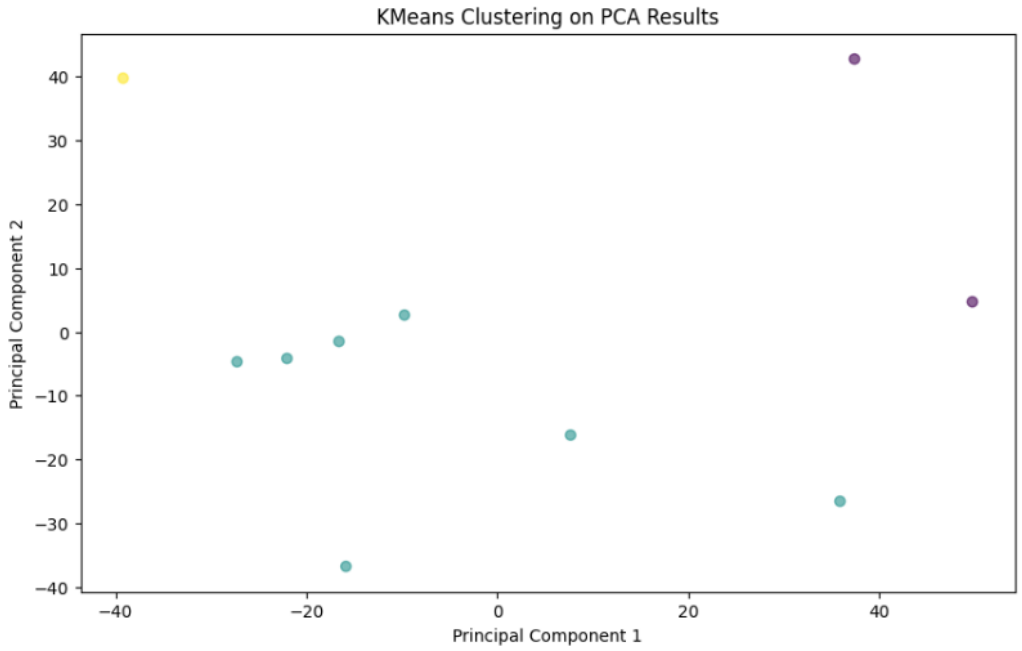
[11]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA

data = pd.read_csv('your_data.csv')
data = data.dropna()

print("Columns in dataset:", data.columns)

plt.figure(figsize=(10, 6))
data['Feature1'].hist(bins=30, color='skyblue', edgecolor='black')  # Replace 'Feature1' with the correct column name
plt.title('Histogram of Feature1')
plt.xlabel('Values')
plt.ylabel('Frequency')
plt.show()

plt.figure(figsize=(10, 6))
data['Feature2'].value_counts().plot(kind='bar', color='orange', edgecolor='black')  # Replace 'Feature2' with the correct column name
plt.title('Bar Plot of Feature2')
plt.xlabel('Categories')
plt.ylabel('Count')
plt.show()

pca = PCA(n_components=2)
pca_result = pca.fit_transform(data.select_dtypes(include=[np.number]))

plt.figure(figsize=(10, 6))
plt.scatter(pca_result[:, 0], pca_result[:, 1], c='green', alpha=0.5)
plt.title('PCA of Data')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()

kmeans = KMeans(n_clusters=3, random_state=42)
data['Cluster'] = kmeans.fit_predict(data.select_dtypes(include=[np.number]))

plt.figure(figsize=(10, 6))
plt.scatter(pca_result[:, 0], pca_result[:, 1], c=data['Cluster'], cmap='viridis', alpha=0.6)
plt.title('KMeans Clustering on PCA Results')
plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.show()
```

```
Columns in dataset: Index(['Feature1', 'Feature2', 'Feature3'], dtype='object')
```




Bar Plot of Feature2
Count
1.0
0.8
0.6
0.4
0.2
0.0
5.627715702647961
58.67471951997892
23.272571457285384
79.91869183127174
15.52322737734827
9.539049027366694
6.149641707380066
34.58218477148631
80.88991173132601
4.375277451193005
Categories


PCA of Data
Principal Component 2
40
30
20
10
0
−10
−20
−30
−40
−40
−20
0
20
40
Principal Component 1



The project applies unsupervised learning techniques to analyze a dataset containing multiple features. The goal is to identify hidden patterns or structures within the data. By using Principal Component Analysis (PCA), the dimensionality of the dataset is reduced, making it easier to visualize and understand. Following this, KMeans clustering is applied to group the data into distinct clusters. These clusters help reveal relationships or similarities between data points that were not immediately obvious.

This project is particularly important because it demonstrates how unsupervised learning can uncover valuable insights in data, even when no prior labels are available. The focus is on understanding the underlying structure of the data and exploring how dimensionality reduction and clustering techniques can be used for pattern recognition.