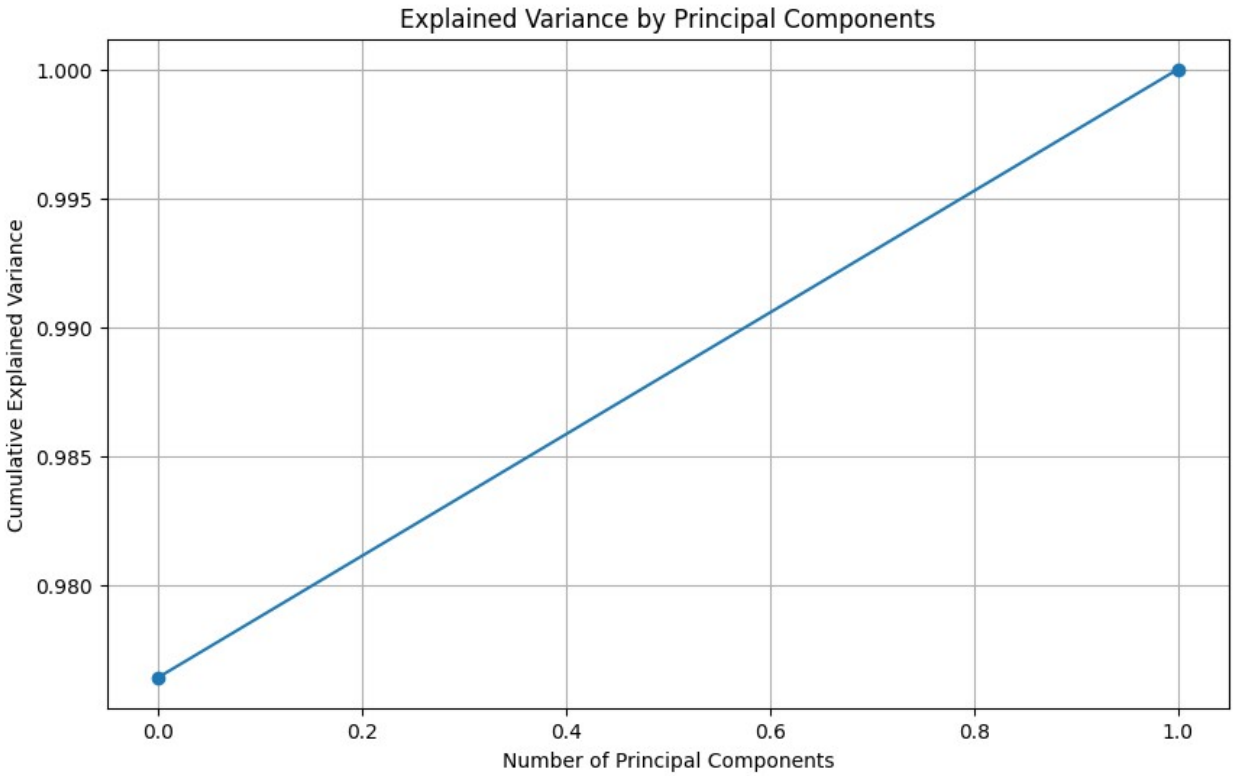
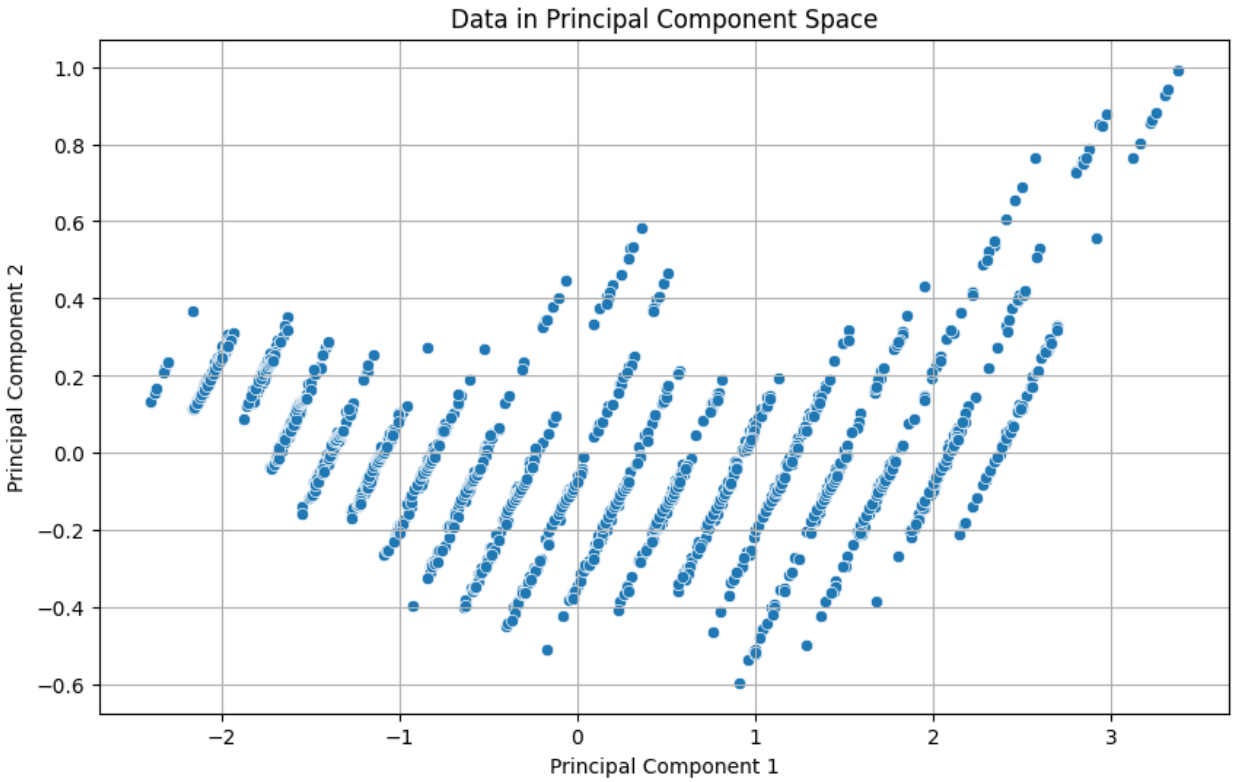
```
import pandas as pd
import numpy as np
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
def load_data(file_path):
    try:
        data = pd.read_csv('/content/rental-price-indexes-september-
2023.csv')
        print(f"\nData loaded successfully from {file_path}")
        print("\nDataset Overview:")
        print(data.info())
        return data
    except FileNotFoundError:
        print(f"File not found: {file_path}")
        return None

# Preprocess the data: Scale the features and drop non-numeric columns
if any
def preprocess_data(data):
    # Select only numeric columns for PCA
    numeric_data = data.select_dtypes(include=[np.number])

    # Scale the features
    scaler = StandardScaler()
    scaled_data = scaler.fit_transform(numeric_data)

    return scaled_data

# Apply PCA
def apply_pca(scaled_data, n_components):
    # Apply PCA transformation
    pca = PCA(n_components=n_components)
    principal_components = pca.fit_transform(scaled_data)

    # Variance explained by each component
    explained_variance = pca.explained_variance_ratio_
    print(f"\nExplained variance by each component:
{explained_variance}")

    return principal_components, explained_variance

# Plot the explained variance
def plot_explained_variance(explained_variance):
    plt.figure(figsize=(10, 6))
    plt.plot(np.cumsum(explained_variance), marker='o')
    plt.xlabel("Number of Principal Components")
```

```
    plt.ylabel("Cumulative Explained Variance")
    plt.title("Explained Variance by Principal Components")
    plt.grid()
    plt.show()

# Main function
def main():
    file_path = input("Enter the path to your dataset CSV file: ")
    data = load_data(file_path)
    if data is None:
        return

    # Preprocess the data
    scaled_data = preprocess_data(data)

    # Ask the user to specify the number of components
    max_components = scaled_data.shape[1]
    n_components = int(input(f"Enter the number of components to keep
(1 to {max_components}): "))

    if n_components < 1 or n_components > max_components:
        print(f"Invalid number of components. Please enter a value
between 1 and {max_components}.")
        return

    # Apply PCA
    principal_components, explained_variance = apply_pca(scaled_data,
n_components)

    # Convert the principal components to a DataFrame
    pca_df = pd.DataFrame(data=principal_components,
columns=[f'PC{i+1}' for i in range(n_components)])
    print("\nFirst five rows of the principal components:")
    print(pca_df.head())

    # Plot cumulative explained variance
    plot_explained_variance(explained_variance)

    # Plot the first two principal components if n_components >= 2
    if n_components >= 2:
        plt.figure(figsize=(10, 6))
        sns.scatterplot(x=pca_df['PC1'], y=pca_df['PC2'])
        plt.xlabel("Principal Component 1")
        plt.ylabel("Principal Component 2")
        plt.title("Data in Principal Component Space")
        plt.grid()
        plt.show()

# Execute the main function
```

```
if __name__ == "__main__":
    main()
```

```
Enter the path to your dataset CSV file: /content/rental-price-indexes-september-2023.csv

Data loaded successfully from /content/rental-price-indexes-september-2023.csv

Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1420 entries, 0 to 1419
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   SER_REF         1420 non-null   object
 1   TIME_REF        1420 non-null   float64
 2   DATA_VAL        1420 non-null   int64
 3   STATUS          1420 non-null   object
 4   UNITS           1420 non-null   object
 5   Subject         1420 non-null   object
 6   Group           1420 non-null   object
 7   Series_title_1  1420 non-null   object
 8   Series_title_2  1420 non-null   object
 9   Series_title_3  1420 non-null   object
dtypes: float64(1), int64(1), object(8)
memory usage: 111.1+ KB
None
Enter the number of components to keep (1 to 2): 2

Explained variance by each component: [0.97639127 0.02360873]

First five rows of the principal components:
        PC1       PC2
0 -2.297608  0.236238
1 -2.312438  0.218514
2 -2.144569  0.128775
3 -2.126844  0.143605
4 -2.099354  0.168201
```


Explained Variance by Principal Components
Cumulative Explained Variance
Number of Principal Components
1.000
0.995
0.990
0.985
0.980
0.0
0.2
0.4
0.6
0.8
1.0


Data in Principal Component Space
Principal Component 2
Principal Component 1
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−2
−1
0
1
2
3