Exercise: Wine Dataset Analysis using K-means and PCA

Objectives:
This exercise will help you understand and implement K-means clustering and Principal Component Analysis (PCA) using the wine dataset.

Dataset: Wine Dataset (available in sklearn.datasets)

Tasks:

1. Data Preparation:
- Import necessary libraries (pandas, matplotlib.pyplot, and sklearn).
- Load the wine dataset using the function `load_wine()` from `sklearn.datasets`. Explore the dataset, check its features, target variables, etc.

2. Data Exploration:
- Perform basic exploratory data analysis. Check the statistical properties of the dataset using pandas `.describe()` method.
- Visualize the distribution of different features in the dataset using suitable plots like histograms or boxplots.

3. Data Preprocessing:
- Standardize the feature matrix. This step is crucial because PCA is affected by the scale of the features. Use `StandardScaler` from `sklearn.preprocessing` to standardize the features to have mean=0 and variance=1.

4. PCA Application:
- Apply PCA to the standardized features. Use `PCA` from `sklearn.decomposition`. Start by considering two principal components.
- Visualize the PCA-transformed data.

5. K-Means Clustering:
- Apply K-means clustering to the PCA-transformed data. Start with a random number of clusters, say k=3. Use `KMeans` from `sklearn.cluster`.
- Visualize the clusters.

6. Choosing Optimal K:
- Use the elbow method to find the optimal number of clusters. Plot the variation of the sum-of-squares within clusters with the number of clusters to visualize the 'elbow'.

7. Presentation of Findings:
- Prepare a summary on your findings. Your summary should include the following points:

  - Introduction of the dataset and the problem.
  - Data preprocessing steps.
  - Application of PCA and reasoning behind the number of chosen components.
  - Application of K-Means clustering and method to choose the optimal number of clusters.
  - Interesting findings in the data.
  - Any real-world applications of the methods applied.

8. Group Discussion:

- We will have a group-wide discussion to share the challenges faced during the exercise, insights gleaned from the dataset, and how these methods can be applied in real-world scenarios.

Hints:

- Don't forget to standardize your data before performing PCA since PCA is scale-dependent.
- For the elbow method, look for the "elbow" in the plot where the within-cluster sum of squares (WCSS) does not decrease significantly with every iteration.

Happy coding!