Python_Lesson6: Python Programming

Lesson Overview:
In this lesson, we will focus on clustering techniques
Clustering using KMeans and Dimension reduction using PCA

Use Case Description:
1. KMeans Clustering and PCA

Programming elements:
KMeans Clustering and Data Analysis

Source Code:
https://umkc.box.com/s/pgqpapgr12wyuj63sgs3162wztpe3q77

In class programming:

1. Apply K means clustering in this data set provided below: https://umkc.box.com/s/a9lzu9qoqfkbhjwk5nz9m6dyybhl1wqy
   - Remove any null values by the mean.
   - Use the elbow method to find a good number of clusters with the KMeans algorithm
2. Calculate the silhouette score for the above clustering
3. Try feature scaling to see if it will improve the Silhouette score
4. Apply PCA on the same dataset.

Data Description can be found in https://umkc.box.com/s/okcxw829g3m5efkljbtwpoft23bmvvnw

***** Bonus points**

1. **Apply kmeans algorithm on the PCA result and report your observation if the score improved or not?**
   a. **You can try different variation like PCA+KMEANS , SCALING+PCA+KMEANS.**
2. **Visualize the clustering of first bonus question**

Evaluation Criteria:
1. Completeness of Features
2. Code Quality (https://en.wikipedia.org/wiki/Best_coding_practices)
3. Time
4. Feedback Submission

Note: *Cheating, plagiarism, disruptive behavior and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL: https://catalog.umkc.edu/special-notices/academic-honesty/*