Python Data Dive Activities Unit 3.3

Complete the provided questions for each of the given datasets.

1. Briefly describe the dataset and its features.

2. Apply Principal Component Analysis (PCA) to your chosen dataset. Explain the meaning of the first few principal components and their variances.

3. Visualize the data using different dimensionality reduction techniques (e.g., PCA, t-SNE). Interpret the resulting clusters or projections.

4. Implement and compare different clustering algorithms (e.g., K-Means, Hierarchical clustering) on your data. Evaluate the performance of each algorithm using appropriate metrics.

5. Discuss the challenges and limitations of unsupervised learning techniques on the chosen dataset. How might these limitations impact the interpretation of results?

Dataset Options:

1. Iris Flower Dataset: archive.ics.uci.edu/dataset/53/iris

2. MNIST Fashion Dataset: kaggle.com/datasets/zalando-research/fashionmnist

3. MovieLens 100K Dataset: kaggle.com/datasets/abhikjha/movielens-100k