Q1:

A pair plot allows us to observe **how two features relate to each other** across the dataset. In the Iris dataset, features like petal_length and petal_width show a **strong positive correlation**, because of the tight upward-slope cluster in their scatter plot. In contrast, sepal_length and sepal_width show a **weaker correlation**, as their scatter plot points are more dispersed.

Setosa seems pretty different compared to other species, particularly in terms of petal width and petal length.

Versicolor and Virginica overlap in many feature combos, especially in sepal length vs. sepal width. This indicates that these two species are extremely similar in nature.

Q3:

In Principal Component Analysis (PCA), the original features are changed into a new collection of uncorrelated features known as principal components. These are arranged according to how much of the variation they can account for.

Applications: PCA can reduce high-dimensional data to two or three dimensions, making it easier to visualize and interpret patterns or clusters.

PCA minimizes the number of features while also maintaining the maximum amount of data variability, which is helpful when working with high-dimensional datasets. The largest variance in the data is captured by the first principal component, the maximum variance orthogonal to the first is captured by the second, and so on.

Q4:

In the PCA of Iris Dataset, Versicolor and Virginica overlap a lot failing to properly distinguish their features. Setosa is distinctly different from the other two species.

t-SNE provides **well-separated clusters** for all three species. In the plot, Setosa (red) is isolated from the rest just like how PCA has done. However, the PCA result failed to differentiate Versicolor and Virginica. But t-SNE provides more tighter and compact clusters compared to PCA.