**CIS 662: Introduction to Machine Learning and Algorithms**

**HW 2**

**OUTPUT:**

**A diagram of a scatter plot

Description automatically generated**

The initial dataset was characterized by its high dimensionality, encompassing a diverse set of attributes. To mitigate the challenges posed by the varying scales of these attributes, a standardization process was implemented. This procedure harmonized all variables to possess a mean of 0 and a standard deviation of 1, ensuring that each attribute contributed equally to the analysis, irrespective of its initial scale.

Following this, Principal Component Analysis (PCA) was applied to the standardized features with the objective of reducing dimensionality to just two principal components. After the PCA transformation, the data was projected into a novel feature space defined by these two principal components. These components represented linear combinations of the original attributes, effectively capturing the most significant variances inherent in the dataset.

**CONCLUSION:**

While certain segments of the data exhibited distinctive patterns following the PCA dimensionality reduction, it's essential to acknowledge the existence of overlaps in a substantial portion of the dataset. This observation suggests that PCA may not be entirely successful in segregating the data based on quartiles and categories.

In summary, it can be concluded that the utilization of PCA for dimensionality reduction may not provide an optimal solution for distinguishing data points categorized into quartiles.

**References:** <https://chat.openai.com/>, https://www.geeksforgeeks.org/principal-component-analysis-pca/