# G. Anand Kumar

012642046

# PCA vs. t-SNE: A Brief Report

## 1. Comparing PCA and t-SNE

PCA and t-SNE are two common techniques for dimensionality reduction, but they work in different ways and serve different purposes.  
  
PCA is a linear method that reduces dimensions while preserving global structure. It helps identify the most important features by finding principal components that capture the highest variance in the data. However, PCA does not effectively separate clusters in our dataset, as seen in the scatter plot where wine quality scores overlapped.  
  
t-SNE, on the other hand, is a nonlinear technique that focuses on local similarities, making it better for clustering. Our t-SNE visualization showed clear groups of wine quality levels, which PCA failed to do. However, t-SNE is computationally expensive and harder to interpret, as the transformed dimensions do not have a clear mathematical meaning.  
  
The choice between PCA and t-SNE depends on the goal. PCA is useful for interpretability and feature selection, while t-SNE is better for clustering and revealing hidden patterns.

## 2. Key Observations from the Visualizations

### PCA Scatter Plot

- The scatter plot did not show clear separations between different wine quality levels.  
- Most points overlapped, meaning PCA did not effectively cluster wine samples.  
- PCA retained only 45.67% of the total variance, leading to significant information loss.

### t-SNE Scatter Plot

- Distinct clusters appeared, grouping similar wine qualities together.  
- Higher-quality wines (scores 7-8) formed separate groups, while lower-quality wines (scores 3-4) clustered together.  
- This confirms that t-SNE is more effective than PCA for visualizing complex structures.

## 3. Conclusion

PCA is faster and easier to interpret, making it useful for dimensionality reduction and feature selection. However, it does not always capture complex relationships in the data. t-SNE, while computationally expensive, reveals hidden patterns and clusters better.  
  
For the Wine Quality dataset, t-SNE provided a more meaningful visualization, whereas PCA was better for understanding variance and reducing dimensions. The choice between PCA and t-SNE depends on whether interpretability (PCA) or clustering (t-SNE) is the main priority.