# Visualisation and Labelling of Clusters:

A screenshot of a computer screen

Description automatically generated

This image depicts a cluster visualization using Principal Component Analysis (PCA). Here's a detailed explanation of its components:

### Components of the Visualization

1. **Scatter Plot**:
   * **Points**: Each dot represents an observation in the dataset.
   * **Colours**: Different colours signify different clusters.
2. **Axes**:
   * **PCA Component 1)**: This represents the first principal component.
   * **Y-axis (PCA Component 2)**: This represents the second principal component.
3. **Colour Bar**:
   * The colour bar on the right side indicates the range of values associated with the clusters, typically representing the cluster index or some other relevant metric.

### Explanation

1. **Principal Component Analysis (PCA)**:
   * PCA is a dimensionality reduction technique used to reduce the number of variables in the dataset while preserving as much variance as possible. It transforms the original variables into a new set of uncorrelated variables called principal components.
   * The first two principal components (PCA Component 1 and PCA Component 2) are used here to project the high-dimensional data onto a 2D plane for visualization purposes.
2. **Clusters**:
   * The plot shows four distinct clusters, each in a different colour.
   * The clusters are spread out in the PCA-transformed space, making it easier to visualize their separation and structure.
3. **Labelling**:
   * The labels for clusters (e.g., different colours) help in identifying which data points belong to which cluster.
   * The colour bar provides a reference to interpret the colours used in the scatter plot.

### Interpretation

* The separation of clusters in the 2D PCA space indicates that the clustering algorithm (e.g., K-means, DBSCAN) has successfully identified groups of similar data points.
* The distinct colours and separation suggest good cluster differentiation, implying that the data points within each cluster are more like each other than to those in other clusters.

### Use Case

This type of visualization is helpful for:

* **Understanding Data Distribution**: It helps in understanding how data points are distributed in the reduced-dimensional space.
* **Evaluating Clustering Results**: It provides a visual assessment of how well the clustering algorithm has performed.
* **Identifying Patterns**: It helps in identifying patterns and relationships within the data.