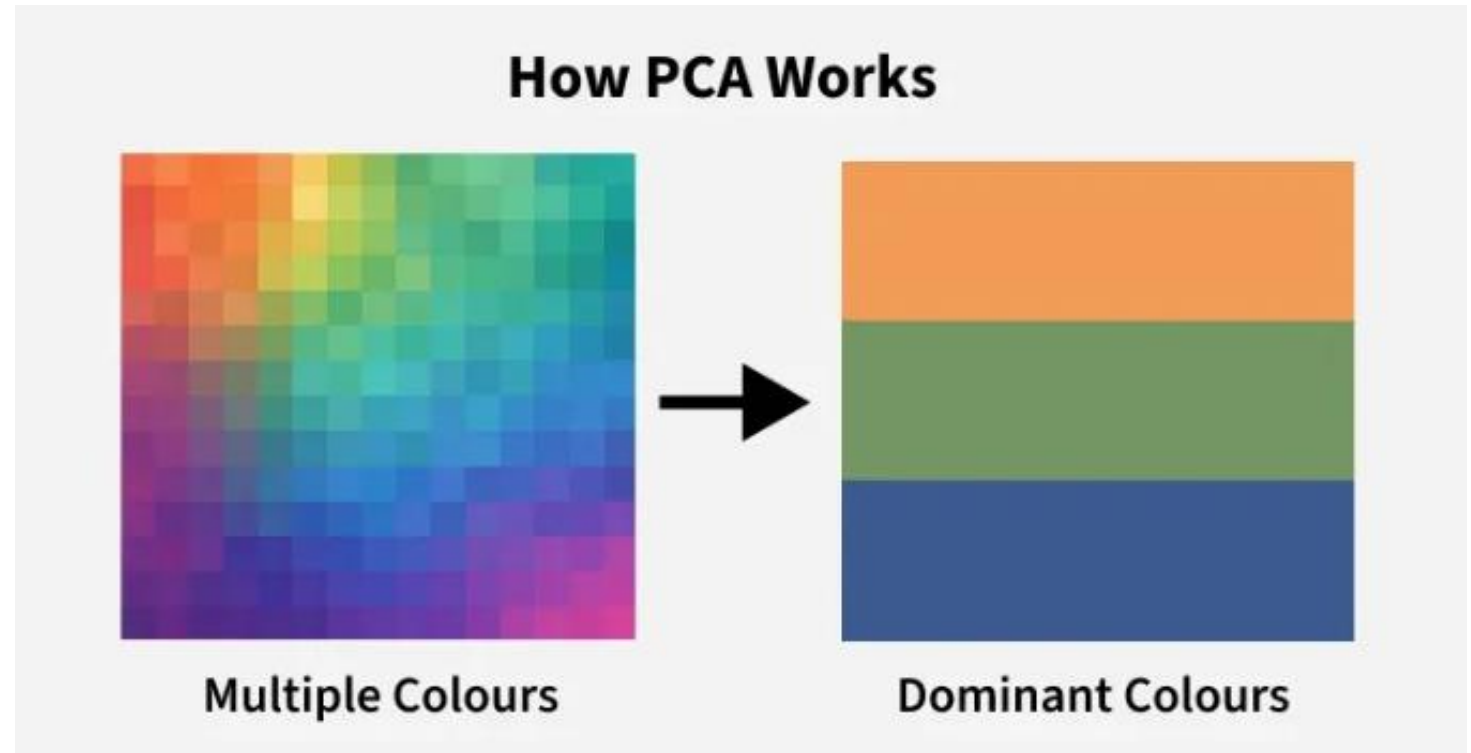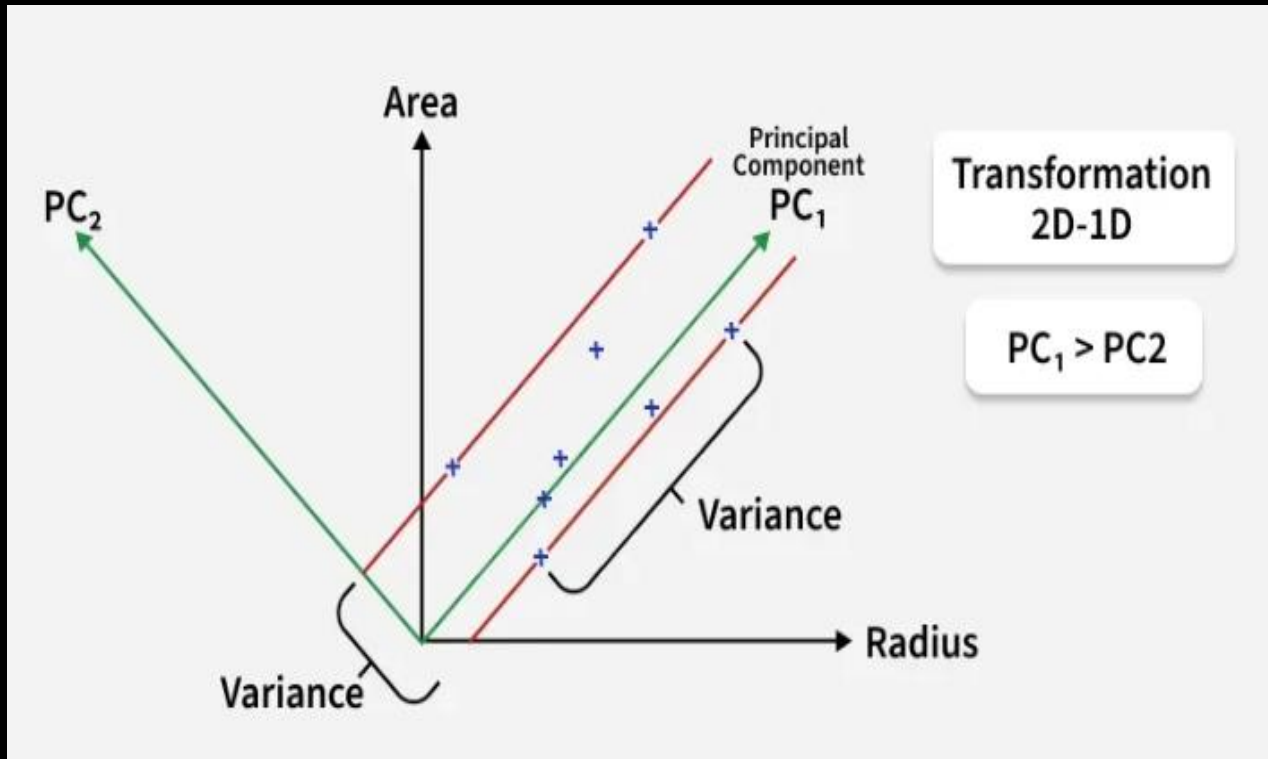# Principle Component Analysis

- **Meaning of PCA**

- PCA is a **dimensionality reduction technique**

- Transforms correlated variables into **uncorrelated principal components**

- Each component captures **maximum variance** in the data

- Components are ordered by **importance (variance explained)**

How PCA Works

Multiple Colours → Dominant Colours

Steps to Perform Principal Component Analysis (PCA)

- **Standardize the Data**
  - Scale features to mean = 0 and standard deviation = 1
- **Compute Covariance Matrix**
  - Measure relationships between variables
- **Calculate Eigenvalues & Eigenvectors**
  - Eigenvectors → principal components
  - Eigenvalues → variance explained
- **Sort Components**
  - Rank components by descending eigenvalues
- **Select Top k Components**
  - Choose components explaining maximum variance
- **Project Data**
  - Transform original data onto selected components

Limitations of Principal Component Analysis (PCA)

Linear Method

Captures only **linear relationships** in data

Loss of Interpretability

Principal components are **combinations of original features**

Hard to assign real-world meaning

Variance ≠ Importance

High variance directions may **not be relevant** for prediction

Sensitive to Scaling

Results depend on **data normalization/standardization**

Affected by Outliers

Outliers can **distort principal components**

Requires Complete Data

Cannot handle **missing values** directly