Unsupervised Learning

Is a type of machine learning where models are trained on data without labeled responses.

Is about discovering general patterns in data

VS

Supervised Learning

Known patterns ==> we use them to do the predictions

Clustering

Grouping similar data points

K-Means clustering

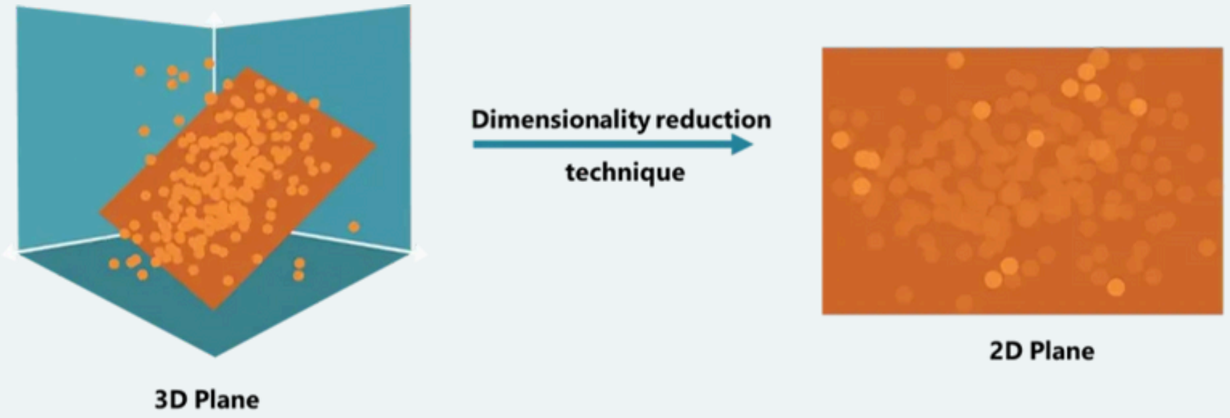
Dimensionality Reduction

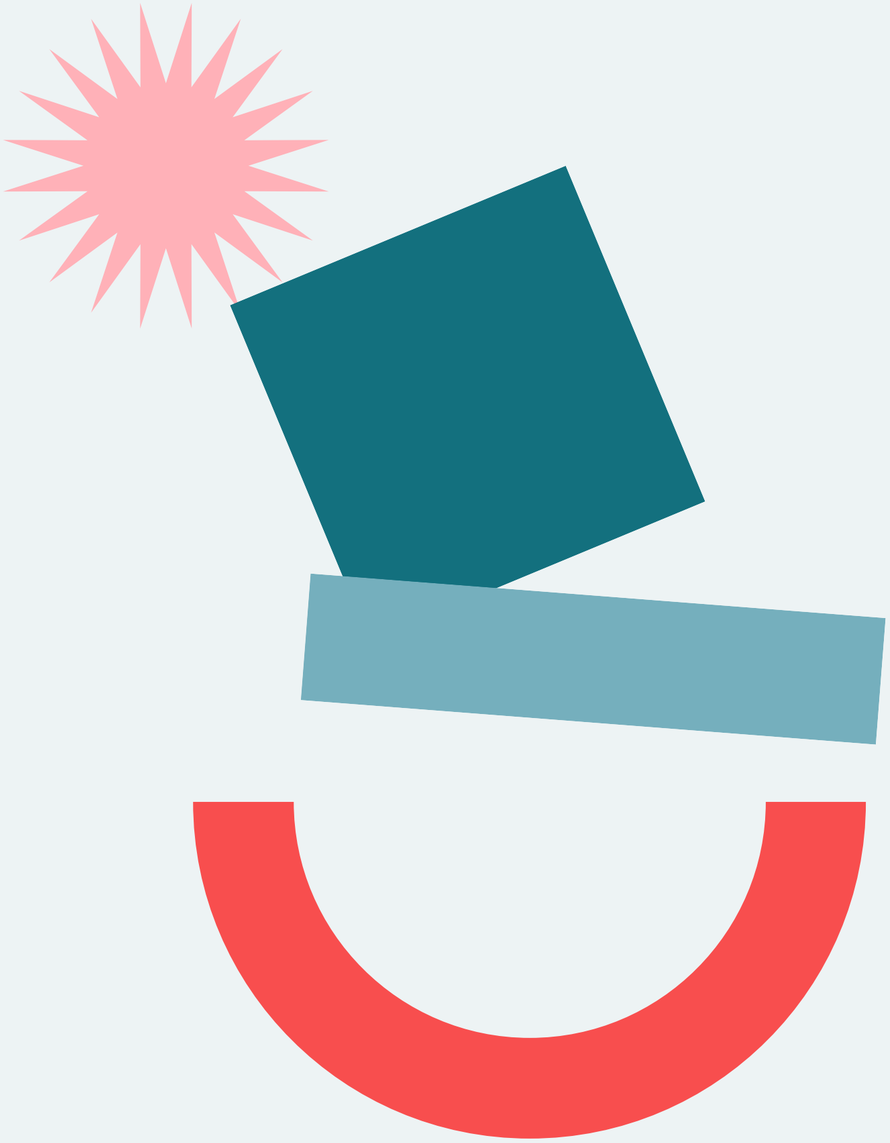
Reducing the number of features while retaining data structure

PCA / T-SNE

PCA

Principal component analysis (**PCA**) is a linear dimensionality reduction technique that can be used to extract information from a high-dimensional space by projecting it into a lower-dimensional sub-space.





t-SNE (t-Distributed Stochastic Neighbor Embedding)

is a technique used to reduce the dimensions of high-dimensional data (like datasets with many features) and visualize it in 2D or 3D space. It helps you understand the structure of complex data by bringing similar data points closer together and dissimilar points further apart.

For example, if you have a dataset with 50 features (dimensions), it's hard to "see" how the data points are organized. This is where t-SNE comes in: it helps reduce the data into two or three dimensions so that we can visualize and understand the structure of the data.

PCA vs t-SNE 🤖🔍

PCA (Principal Component Analysis) is like a smart shortcut 🛣️ that finds the most important directions 🧭 in high-dimensional data and flattens it. It keeps the biggest patterns and helps us see the big picture 🖼️. But... it's linear ⚖️ and may miss the small details 💡!

t-SNE (t-Distributed Stochastic Neighbor Embedding) is the artist 🎨 that turns high-dimensional chaos into beautiful 2D/3D plots 🏞️. It's all about keeping similar things together 👫, but it's non-linear 🌪️ and focuses on local patterns. It's perfect for seeing clusters 💬 but can get a bit messy if you zoom out 🔭.

In short:

- PCA: 📊 Simplifies, keeps the big trends.
- t-SNE: 🌟 Focuses on local structure, creates beautiful (visualizations).