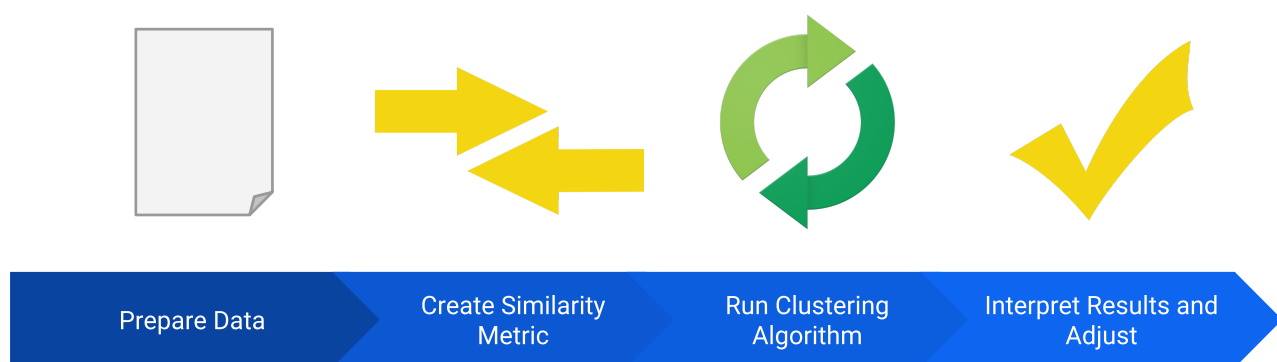# Clustering Workflow 🔖

To cluster your data, you'll follow these steps:

1. Prepare data.

2. Create similarity metric.

3. Run clustering algorithm.

4. Interpret results and adjust your clustering.

This page briefly introduces the steps. We'll go into depth in subsequent sections.

Prepare Data

As with any ML problem, you must normalize, scale, and transform feature data. While clustering however, you must additionally ensure that the prepared data lets you accurately calculate the similarity between examples. The next sections discuss this consideration.

v: For a review of data transformation see Introduction to Transforming Data (hine-learning/data-prep/transform/introduction) from the Data Preparation and Feature Engineering for ne Learning course.

Create Similarity Metric

Before a clustering algorithm can group data, it needs to know how similar pairs of examples are. You quantify the similarity between examples by creating a similarity metric.

Creating a similarity metric requires you to carefully understand your data and how to derive similarity from your features.

Run Clustering Algorithm

A clustering algorithm uses the similarity metric to cluster data. This course focuses on k-means.

Interpret Results and Adjust

Checking the quality of your clustering output is iterative and exploratory because clustering lacks "truth" that can verify the output. You verify the result against expectations at the cluster-level and the example-level. Improving the result requires iteratively experimenting with the previous steps to see how they affect the clustering.